# **BMDExpress2 EcoHTTr User Protocol**

 Authors: Monique Hazemi and Peter Schumann

Last updated: May 2024

# **Overview**

BMDExpress2 is a software tool designed to derive Benchmark Dose (BMD) values from toxicogenomic data, which are widely used in risk assessment. In this protocol, the input data is assumed to be transcriptomic count data obtained by following the "*RNA Sequencing Analysis*" protocol. However, BMDExpress2 can handle various types of toxicogenomic data, including microarray data, targeted gene quantification (e.g., TempO-seq), and other "off-label" data sources.

The Benchmark Dose (BMD) is defined as the dose or concentration that produces a predetermined change in the response rate of an adverse effect, known as the Benchmark Response (BMR). The BMR value is user-defined and can be set as a percentage of the control value, a factor of the standard deviation, or other criteria.

Our goal, within ecotoxicology, is to estimate a transcriptional point of departure (tPOD) for a particular chemical or exposure scenario using the gene-level BMDs calculated from using BMDExpress software. A tPOD is the estimated dose or concentration of a chemical at which transcriptional changes (alterations in gene expression) in a biological system or organism become significant or diverge from the control or baseline state.

**Key Features of BMDExpress2**:

* Integration of benchmark dose calculations with functional classification analyses based on Gene Ontology (GO), Signaling Pathways (REACTOME), or custom categories provided by the user.
* User-friendly workflow with visualizations of results and numerous filtering options.
* The BMD analysis curve-fitting feature is based on the EPA's Benchmark Dose Software (BMDS), which is widely used in human health risk assessments, ensuring reliable and standardized calculations.

By combining toxicogenomic data analysis with benchmark dose modeling and functional enrichment, BMDExpress2 provides a powerful tool for researchers and risk assessors to evaluate the potential adverse effects of compounds and identify relevant biological pathways and processes.

**Note**: A newer version, ***BMDExpress3***, is currently under active development. While it offers additional features and improvements, BMDExpress2 has undergone more extensive peer review and validation. This protocol is applicable to both BMDExpress2 and BMDExpress3, as the core functionality and workflow remain similar.

Link to the "RNA Sequencing Analysis" protocol:



## **User Manual and Tutorial Information:**

* [*National Toxicology Program Report (2018)*](https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/rr05_508.pdf?utm_source=direct&utm_medium=prod&utm_campaign=ntpgolinks&utm_term=rr05)

- Details recommended parameters to use when running a dataset through BMDExpress. We use these guidelines to direct our analyses.

* [*GitHub Tutorial and Useful Software Information*](https://github.com/auerbachs/BMDExpress-2/wiki#import-dose-response-data)

*-* Includes what each test/analysis parameter means.

* [*YouTube video tutorial notes*](https://www.youtube.com/playlist?app=desktop&list=PLX2Rd5DjtiTezKzk9Inf5nn5GWZ8TRBzd)

# **BMDExpress Protocol**

## **1 Loading count data**

1. Open BMDExpress2 (or BMDExpress3)
2. (*Optional*) Update annotations

**Note**: This step typically does not apply to non-model organisms, but it is a good idea to check in case the software was updated with annotations for the species you are analyzing.

* 1. File 🡪 “Update Annotations”
	2. Select your platform from the list by checking the box in the “Select” column.
		1. You can sort these options by various parameters such as species, provider, or name.
	3. Click “Update”
	4. Click “Done” once it is finished.
1. Load in the count matrix file generated using the "RNA Sequencing Analysis" protocol. This should be a .txt file ending with “BMDExpress\_counts”. When loading this file into Excel, it should be formatted like this:



**Additional notes on data file formatting**

* + - The cell A1, should always be left blank.
		- Depending on the species, sequencing platform, experimental design, etc., there may be 10s of thousands of rows representing distinct gene/transcript features.
		- Dose treatment columns can be out of numerical order (e.g., Dose: 0,0,4.6,2.6,0,47.1). The software will recognize this and sort appropriately.
1. To load in the count matrix file, go to “File” 🡪 “Import” 🡪 “Expression Data”
2. Select the folder and file name of your data in pop up dialog box.
	1. You can batch load multiple files by pressing “Shift + click” and then hit “Open”.
	2. From the “Platform chooser”, click the drop down and select “Generic” (found at the bottom of the list)
	3. From the “Log Transformation chooser”, select “Base2" (since we are working with log2 transformed data) then click “Ok”
	4. You should now see that the data is listed under the side navigation panel under “Expression Data”:



* 1. Click on the check box next to one of the file names to verify that the data was imported.
		1. You may want to record the number of features that were detected shown by “Total Items” (see [section 4](#_4_Recording_Data)):



**Note**: If you imported the wrong data file, you could right click on the file name in the left-hand side menu and select “Remove”.

**Saving your project:**

BMDExpress is not bug-free! It is a good idea to save your project after each step of this protocol or just as often as possible. To save your project for the first time, go to “File” 🡪 “Save Project As” and select a file location to save the project to. Subsequent saves can be done by going to “File” 🡪 “Save Project” for faster saves. Throughout this protocol, you will see reminders to “**SAVE PROJECT!”**

## **2 Prefiltering**

### **2.1 One-way ANOVA**

A one-way analysis of variance (ANOVA) is used to identify gene features that show significant differences in expression levels across the control and treatment groups. For each gene, the one-way ANOVA tests if the mean expression levels are equal across *all groups*. Genes with a p-value less than the chosen significance level are considered responsive to the chemical treatment, while those with a higher p-value are deemed non-responsive and can be filtered out from further analyses. This is not looking at dose responsiveness, rather, it is to determine if the gene feature is responsive to the chemical at all.

**Important**: If the ANOVA test for a count matrix results in 0 responsive features, then typically we do not proceed with further analysis for that sample set. However, this is not a strict guidance.

1. Highlight the data file name that you want to analyze by clicking on it in the side navigation panel under “Expression Data”.
	1. You can analyze multiple files at once and select them by using “Ctrl + Click” to select them one-by-one or “Shift + Click” to select them down the entire list. BMDExpress will analyze each file independently.
2. Run the ANOVA test by going to Tools 🡪 Prefilter 🡪 One-way ANOVA
3. Select the following recommended testing parameters:
	* + - P-Value cutoff = 0.05
			- Check the box for “Multiple Testing Correction”
			- Check the box for “Filter Out Control Genes”
			- Uncheck the box for “Use Fold Change Filter”
			- NOTEL/LOTEL Determination = T-test
			- NOTEL/LOTEL P-Value = 0.05
			- NOTEL/LOTEL Fold Change Value = 2.0
			- Number of threads = *Depends on your computer*!
				1. To determine this (assuming you are using a Windows OS), click “Ctrl + Shift + Esc” on you keyboard to bring up the Task Manager.
				2. Navigate to the “Performance” page:



* + - * 1. In the bottom right of this window, you will see how many “cores” and “logical processors” your computer has:

****

Ideally, if you don’t want to overburden your computer, you should select half of the number of logical processors available. In this case, we would choose 4 for the “Number of threads”.

Here is what this input selection looks like in BMDExpress2:



1. Click “Save Settings” so that you can reuse these parameters, then click “Start”.
2. When the ANOVA test is complete, the results will appear in the side navigation panel under “One-way ANOVA”. Use the drop-down menu to navigate to it:

 🡪

1. Check the box next to the file’s name to view the results.
	* 1. In the results table, you will find the Probe/Gene ID, the ANOVA statistical parameters, fold change values, fold change values at each dose level, and more.
		2. In charts area, you will find several different visualizations. Volcano plots are set as default. You can download, enlarge, and edit the plots by clicking on the various icons above each: 
		3. The testing parameters you selected are detailed in the bottom left pane.
		4. You may want to record the number of features that passed the test shown by “Total Items” as in [section 1](#_1_Loading_count) step 7a

**SAVE PROJECT!**

### **2.2 Williams Trend Test**

The Williams Trend Test is used to identify genes or probes that exhibit a significant monotonic trend (an increasing or decreasing *dose-response relationship*) in expression levels across ordered dose groups. Genes with a p-value less than the chosen significance level are considered dose-responsive, while those with a higher p-value are deemed non-responsive and can be filtered out.

1. To perform this test, highlight the data file name that you want to analyze by clicking on it in the side navigation panel under “Expression Data”.
	1. You can analyze multiple files at once and select them by using “Ctrl + Click” to select them one-by-one or “Shift + Click” to select them down the entire list.
2. Then go to “Tools” 🡪 “Prefilter” 🡪 “Williams Trend Test”.
3. Select the following recommended testing parameters:
	* + - P-value cutoff = 0.05
			- Number of permutations = 100
			- Uncheck the box for Multiple Testing Correction
			- Check the box for Filter Out Control Genes
			- Check the box for Use Fold Change Filter
			- Fold change value = 2.0
			- NOTEL/LOTEL Test = T-test
			- NOTEL/LOTEL p-value = 0.05
			- NOTEL/LOTEL fold change value = 2.0
			- Number of threads = see [section 2.1](#_2.1_One-way_ANOVA) step 3

Here is what this input selection looks like in BMDExpress2:



1. When the Williams Trend Test is complete, the results will appear in the side navigation panel under “Williams Trend Test”. Use the drop-down menu to navigate to it.
	* 1. You may want to record the number of features that passed the test shown by “Total Items” as in [section 1](#_1_Loading_count) step 7a. This is the number of *concentration responsive genes* (CRGs).
		2. The plots can be downloaded or modified as described for the one-way ANOVA results, at the user’s discretion.

**SAVE PROJECT!**

## **3 Benchmark Dose Analysis**

BMD values are calculated for each concentration responsive gene determined by the Williams Trend Test. This involves fitting the normalized gene expression data to various dose-response models. The best-fitting model is then used to derive the BMD, which is the dose corresponding to a benchmark response level. For detailed statistical methods underlying this approach, please refer to [Yang et al., 2007](https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-8-387).

### **3.1 Dose-response curve fitting**

1. Highlight the Williams Trend Test result that you want to analyze by clicking on it in the side navigation panel under “Williams Trend Test”:

 🡪

* 1. You can analyze multiple files at once and select them by using “Ctrl + Click” to select them one-by-one or “Shift + Click” to select them down the entire list.
1. To initialize dose-response model fitting, got to “Tools” 🡪 “Benchmark Dose Analysis” 🡪 “EPA BMDS Models”.
2. Select the following recommended testing parameters:
	1. First, select the models that you want to perform curve fitting with under “Continuous Models”

**Note**: The more models you select, the longer it will take to perform the analysis. Depending on how many CRGs you have, the total time for analysis could be hours (or days!). Additionally, there are different models available in BMDExpress2 and 3. Typically, we use all models except for “Poly3” and “Poly4”.

* 1. Select the remaining parameters:

**Note**: Not all these parameters apply to BMDExpress3. Refer to the figures below for guidance on selecting parameters for that version.

* + - * Max iterations = 250
			* Confidence level = 0.95
			* Check the box for “Constant Variance”
			* BMR Type = Standard Deviation
			* BMR Factor = 1 SD
			* Restrict Power = No restriction
			* BMDL and BMDU = Compute and utilize in best model selection
			* Best Poly Model test = Nested chi square
			* Check the box for “Flag Hill Model with ‘k’ parameter <”
				+ Select “1/3 of lowest positive dose” from the drop-down menu
			* Best model selection with flagged hill model = Select next best model with P-Value > 0.05
			* P-Value Cutoff = 0.05
			* Number of threads = see [section 2.1](#_2.1_One-way_ANOVA) step 3
			* Model execution timeout (secs) = 600 (default)

Here is what this input selection looks like in **BMDExpress2**:



Here is what this input selection looks like in **BMDExpress3**:



**Note**: The *BMDL and BMDU* parameter selected is “Compute and utilize BMD, BMDL and BMDU in best model selection”

**SAVE PROJECT!**

### **3.2 Results visualizations**

When BMDExpress has finished running the BMD analysis, you can access the results through the side navigation panel under “Benchmark Dose Analyses”. Check the box next to the result name to view it.

* + In the results table, you will find the Probe ID, Entrez gene IDs ("Genes"), Gene symbols, the model parameters, etc.
		- If you want to see an individual probe and its model fit, select the desired probe by either double clicking on it in the table or selecting it and pressing “Enter” on your keyboard. A dialog box will come up that includes a graph showing curve fit and the model equation and BMDL, BMD, BMDU, Fit P and AIC values:



↓



* + - You can view the dose-response curves for different genes without closing out of the plot window by clicking a different gene from results table. Or you can select desired probe ID in the ID section within plot window.
	+ In charts area, you can find several different visualizations:
		- Pie chart that shows distribution of best fit models
		- Best BMD to BMDL scatterplot
			* You should see a roughly linear (1:1) relationship here. If you don't, that means your fits aren't so great.
	+ **Histogram distributions of the best BMD**
		- As we typically define a tPOD as the 10th percentile of the BMD values, this is a particularly useful plot to download and save to understand the distributions of the gene BMDs within a treatment group.
		- Other types of visualizations can be pulled up using the Select Chart View drop-down menu
		- You can pull up a spreadsheet view of all the probe IDs that fitted to Hill models, for example, by right clicking on your desired Benchmark dose analysis result 🡪 Spreadsheet View.

### **3.3 Filtering BMD results**

Before deriving a tPOD, we need to filter the BMD results to retain only the most relevant BMD estimates.

* 1. On the right-hand side of the BMDExpress window, there should be a filter pane like this:



* If not, click the “Show Filter” button just above the visualizations pane:



* 1. From the Add Filter drop-down menu, select the following filters and criteria:
		+ - Best BMD 🡪 “between” 🡪 left box: lowest dose used in this study divided by 10 🡪 right box: the highest dose used in this study
				* **Note**: You can find this information in the lower right information pane:



* + - * Best BMDU/BMDL < 40
			* Best fitPValue > 0.1
		- The results (table and visualizations) will auto-update as you select each filter. If not, check the “Apply Filter box” above the visualizations pane:




### **3.4 Exporting BMD results**

You will need to export the final BMD analysis results as a .txt file for use in the “analyze\_bmd()” function found in the “BMD\_summary.R” script. See the "RNA Sequencing Analysis" protocol, step 4.4.4 for details.

1. Select “Export Filtered Data” above the visualizations pane:



* 1. Navigate to a folder that will store all the exported BMD results for a particular project 🡪 “Save”.
		+ Typically, a folder called “all\_bmdexpress\_data” is created within a project folder to hold the exported data, the project file, and any downloaded visualizations.
		+ As for a file name, we suggest using the default BMDExpress format, as it retains the most specificity and ends with “\_BMD\_filtered.txt”, which is a *required* file name ending to perform the tPOD calculations using the “analyze\_bmd()” function.

## **4 Recording Data**

Below is a recommended list of *columns* for a data spreadsheet for saving key data from the BMD analysis:

|  |
| --- |
| **Chemical Name** |
| **Sample Replicate/Gene Set** |
| **Analysis Date** |
| **Analyzer** |
| **Total Features** |
| **Passed ANOVA?** |
| **ANOVA Features** |
| **CRGs** |
| **Highest Conc. (uM)** |
| **Lowest Conc. (uM)** |
| **Filtered CRGs** |
| **Median BMD (uM)** |
| **Median BMDL (uM)** |
| **Median BMDU (uM)** |
| **BMDL 10th percentile** |
| **BMDU 10th percentile** |
| **tPOD (uM)** |

**Note**: The concentration/dose units are shown as uM, but you can use whatever units make the most sense for your dataset.