DoseRider: User Guide

Pablo Monfort Lanzas*

March 30, 2025

*Email: pablo.monfort@i-med.ac.at

Contents

1 Overview of DoseRider and its Purpose

DoseRider is a web-based and R package tool designed to facilitate researchers in analyzing multi-omics dose-response data at the pathway level. The platform provides a robust and flexible framework for modeling both linear and non-linear dose-response relationships using generalized linear mixed models (GLMM) with cubic splines. The goal is to capture complex biological responses to various doses of compounds or chemicals, allowing users to compute critical metrics like Trend Change Doses (TCD) and Benchmark Doses (BMD) for entire pathways.

In toxicogenomics, dose-response modeling typically focuses on individual genes; however, **DoseRider** extends this analysis to the pathway level, providing a more holistic and mechanistic understanding of how specific gene sets respond to different concentrations of exposure compounds. By integrating custom and pre-annotated gene sets, **DoseRider** uncovers essential molecular patterns, enabling researchers to make informed decisions in the context of toxicological risk assessment and regulatory guidelines.

The purpose of **DoseRider** is to offer researchers an intuitive and comprehensive platform that streamlines the analysis of dose-response data across various omics layers, such as transcriptomics, proteomics, and metabolomics. With automated model selection, pathway-specific BMD calculations, and advanced visualizations, the tool enhances the depth and accuracy of dose-response studies, thereby improving the evaluation of chemical safety.

DoseRider supports data for a variety of species, including *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Danio rerio*, *Drosophila melanogaster*, and *Caenorhabditis elegans*. The web application can be accessed at http://doserider.icbi.at, while the R package is available at https://github.com/icbi-lab/doserider.

2 DoseRider's Workflow

2.1 Workflow overview

DoseRider workflow consists of four key stages, as shown in Figure 1:

1. **Data Preprocessing:** The user uploads dose-response datasets, including raw counts from RNA-seq experiments or intensity values from microarray or metabolite experiments. Dose information and metadata, such as doses administered and sample identifiers, must also be provided. The data is then log-transformed if required for wide dose ranges.
2. **Pathway Modeling:** DoseRider utilizes generalized linear mixed models (GLMMs) to model the dose-response relationships for entire pathways. Each pathway is evaluated using various models, including null models, linear models, and non-linear models with cubic splines. The tool applies random effects at the gene level to account for inter-gene variability.
3. **Model Selection and Validation:** For each pathway, **DoseRider** selects the best-fitting model based on Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and significance testing (p-value ¡ 0.05). The selected models are used to predict dose-response curves and identify key metrics, including TCD and BMD.
4. **Visualization and Reporting:**The results are compiled into pathway results tables, and users can visualize dose-response relationships through heatmaps, dose-response curves, and random effects plots. All results can be downloaded in various formats for further analysis.

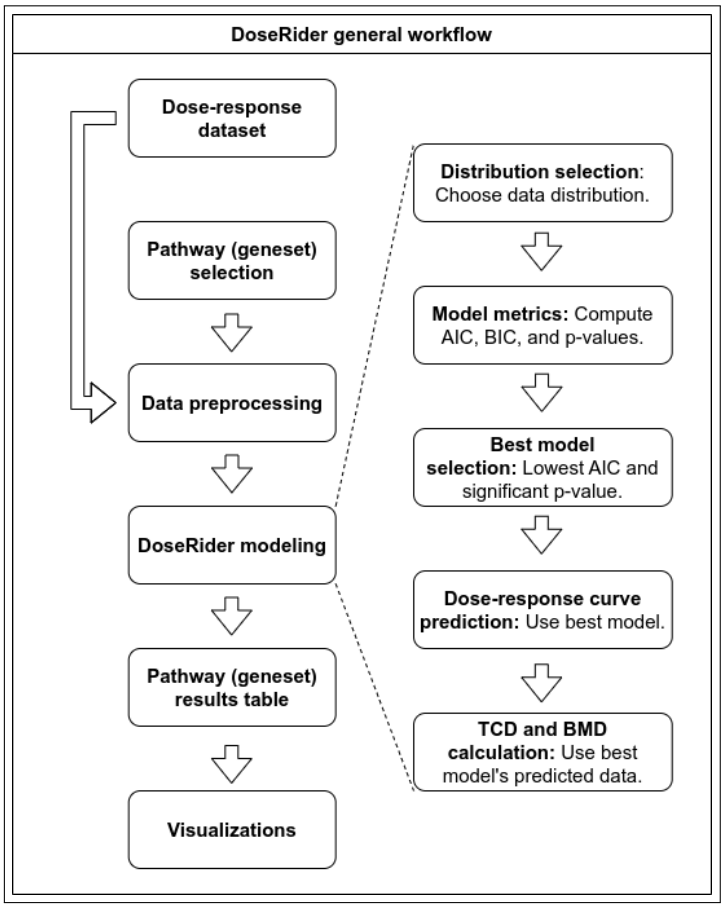

Figure 1: **DoseRider workflow.**

2.2 Data Preprocessing

Data preprocessing is a crucial initial step in the DoseRider workflow, as it ensures that the input data is formatted correctly and is ready for dose-response analysis. DoseRider accepts a variety of multi-omics data types, such as RNA-seq, microarray, proteomics, and metabolomics data, which must be processed and normalized prior to model fitting. The preprocessing phase involves the following steps:

2.2.1 Input Data Requirements

DoseRider expects two primary input files:

- **Expression Data:** This file contains raw or normalized counts for gene expression (in the case of RNA-seq) or intensity values (for microarray, proteomics, or metabolomics experiments). Each row corresponds to a gene or feature, and each column represents a sample or replicate. The data must be in tabular format, typically as a CSV or TSV file.
- **Metadata:** This file includes information about the experimental conditions, such as dose levels, sample identifiers, and additional experimental metadata (e.g., exposure time). The metadata is linked to the expression data through sample identifiers, ensuring that each sample has corresponding dose information.

If survival data is included, DoseRider can also calculate half-maximal inhibitory concentration (IC50) values. This requires additional columns in the metadata file specifying survival rates or other phenotypic measures.

2.2.2 Log Transformation and Normalization

In dose-response experiments, doses are often distributed across several orders of magnitude. For such datasets, DoseRider recommends applying log transformation to the dose values to stabilize variance and improve the interpretability of the model. The user can choose to log-transform doses directly within the web interface.

2.2.3 Gene Set and Feature Filtering

DoseRider includes a gene set filtering step to ensure that only relevant features are included in the dose-response modeling process. This step can be customized to the user's needs:

- **Minimum and Maximum Gene Set Size:** Users can set thresholds for the minimum and maximum size of gene sets to be included in the analysis. By default, gene sets with 20 to 200 genes are considered, though this can be adjusted based on the complexity of the dataset.

DoseRider includes flexible filtering options to ensure that only informative and coherent gene sets are included in dose-response modeling. These steps help exclude pathways with low signal, excessive noise, or antagonistic gene behavior. All thresholds can be customized by the user.

- **Minimum and Maximum Gene Set Size:** Gene sets must contain a user-defined number of genes (default: between 20 and 200) to be eligible for analysis.
- **Low Variance Filter:** Pathways are excluded if a large proportion of genes have low variance (below the 10th percentile across all genes), indicating limited dynamic response.
- **Negative Correlation Filter:** Pathways are removed if more than 50% of gene pairs are negatively correlated, suggesting conflicting expression patterns.
- **Principal Component Filter:** Gene sets are excluded if the first principal component explains over 70% of total variance, typically reflecting opposing gene trends that obscure a common response.
- **Automatic Clustering of Divergent Patterns:** For pathways with heterogeneous responses, clustering is applied automatically to identify gene subgroups with distinct dose-response behaviors.

2.3 Pathway Selection

DoseRider employs a robust algorithm to determine the most responsive pathways for dose-response analysis. For each gene set, the tool curates a collection of pathways based on the user-specified dataset and the available annotations in the Molecular Signatures Database (MSigDB), Immune-Related Gene Signatures (MIO), or any custom GMT file uploaded by the user. The pathways are initially selected based on size, with the default range set between 20 and 200 genes. This ensures that gene sets are sufficiently representative while minimizing computational load.

The tool further refines pathway selection using Principal Component Analysis (PCA). Pathways where the first principal component (PC1) explains over 70% of the variance are prioritized. Additionally, pathways with an antagonist score below a specified threshold are excluded from analysis, as these exhibit significant opposing gene expression patterns.

Here is a refined version of the **Modeling Dose-Response Relationships** section, rewritten to align with the content and technical details found in the provided **DoseRider** documents:

2.4 Modeling Dose-Response Relationships

DoseRider employs a sophisticated framework of generalized linear mixed models (GLMMs) to capture dose-response relationships at the pathway level. This approach allows researchers to explore both linear and non-linear dose-response behaviors across various omics datasets, such as transcriptomics, proteomics, and metabolomics, while accommodating the inherent variability between individual genes within a pathway.

The modeling process in **DoseRider** involves several stages:

2.4.1 Model Selection

For each gene set or pathway, **DoseRider** evaluates several models to describe the relationship between dose and gene expression. These models include:

- **Null Model:** Assumes no relationship between dose and gene expression, serving as a baseline.
- **Linear Model:** Assumes a direct, linear relationship between dose and gene expression, with dose as the fixed effect.
- **Non-Linear Fixed Model:** Utilizes cubic splines to model dose-response relationships, allowing for non-linear trends. This model is essential when the biological response is not simply proportional to the dose.
- **Non-Linear Mixed Model:** Extends the non-linear model by incorporating random effects, specifically accounting for variability across genes within the same pathway. Random effects in this model allow for gene-specific dose-response curves, enhancing the model's flexibility and accuracy.

DoseRider dynamically selects the best model for each pathway based on the lowest corrected Akaike Information Criterion (AIC) and statistical significance (p-value <0.05) when compared to the null model. The use of AIC ensures a balance between model complexity and goodness-of-fit, avoiding overfitting while capturing critical dose-response patterns.

2.4.2 Smoothing Dose-Response Curves

Once the best model is selected, **DoseRider** generates smoothed dose-response curves using the Best Linear Unbiased Predictor (BLUP). These smoothed curves provide a more intuitive representation of gene expression changes across varying doses, helping to visualize complex non-linear relationships that are typical in toxicogenomics.

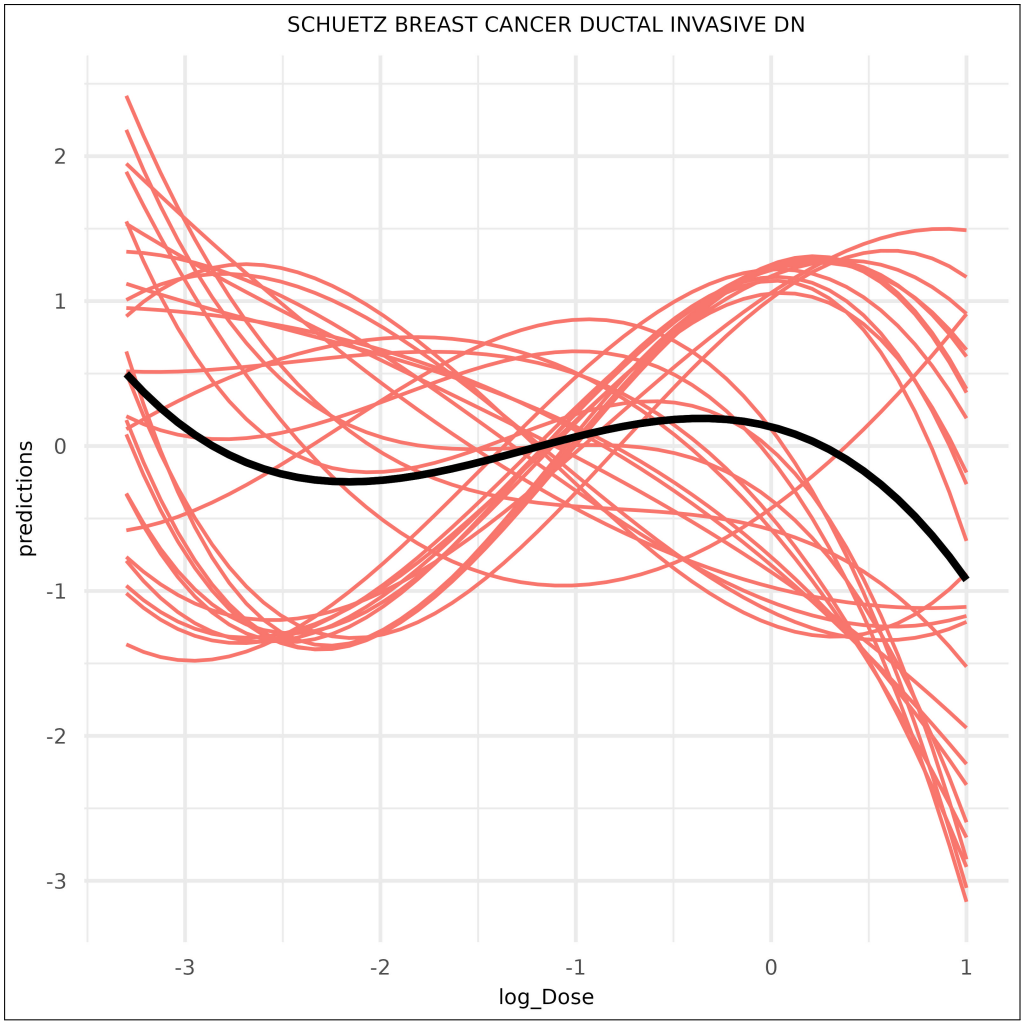

Figure 2: **Dose-Response Modeling in DoseRider.** Non-linear dose-response models using cubic splines with gene-specific random effects allow for accurate representation of gene expression changes across dose levels.

2.4.3 Benchmark Dose (BMD) Calculation

A critical feature of **DoseRider** is the calculation of the **Benchmark Dose (BMD)** for each pathway, a key metric used in toxicological risk assessment. The BMD represents the dose at which a significant change in gene expression occurs, relative to a control dose. **DoseRider** computes BMD values based on the smoothed dose-response curves using the BMD-zSD approach. This method estimates the dose where gene expression deviates by a defined number of standard deviations from the baseline expression level, offering a robust and reliable measure for setting regulatory thresholds.

For pathways with biphasic dose-response behaviors, where gene expression increases at low doses and decreases at higher doses (or vice versa), **DoseRider** calculates the BMD at the lowest dose where the benchmark response is observed. This allows for the identification of early responses to toxicants, which are critical for assessing low-dose exposures.

2.5 Trend Change Dose (TCD) Calculation

In addition to the Benchmark Dose (BMD), DoseRider introduces the concept of Trend Change Doses (TCDs) to capture inflection points in complex dose-response curves. While BMD estimates a threshold for statistically significant responses, TCDs detect shifts in the curve's direction—providing earlier and more sensitive indicators of biological activity.

Mathematical Basis

TCDs re identified using the first and second derivatives of the predicted dose-response function, typically modeled using spline-based mixed models. A TCD corresponds to a local minimum, maximum, or an inflection point in the curve.

Bootstrap-Based Confidence Intervals

To quantify uncertainty, DoseRider computes 95% confidence intervals for TCDs via bootstrap resampling. This approach reflects the variability in trend direction estimates due to noise or model instability and provides robust bounds around each TCD.

Interpretation

TCDs offer a complementary perspective to BMDs by detecting early changes in gene or pathway behavior that may signal adaptive responses, homeostatic mechanisms, or transition states. These shifts do not necessarily indicate adverse effects but may highlight critical stages in a compound's mode of action.

Output Visualizations

DoseRider includes the following visualizations to support TCD interpretation:

- **TCD Density and Peaks:** Global distribution of TCDs across pathways to identify dose regions with frequent trend reversals.
- **TCD Confidence Intervals:** Estimated uncertainty ranges for each TCD based on bootstrap analysis.

2.5.1 Random Effects and Variability Modeling

To account for the natural variability in gene responses within a pathway, **DoseRider** incorporates gene-specific random effects into the model. This allows for greater flexibility and accuracy in modeling pathways with heterogeneous gene expression patterns. Random effects are particularly useful in toxicogenomics, where different genes within a pathway may respond to doses at different rates or intensities.

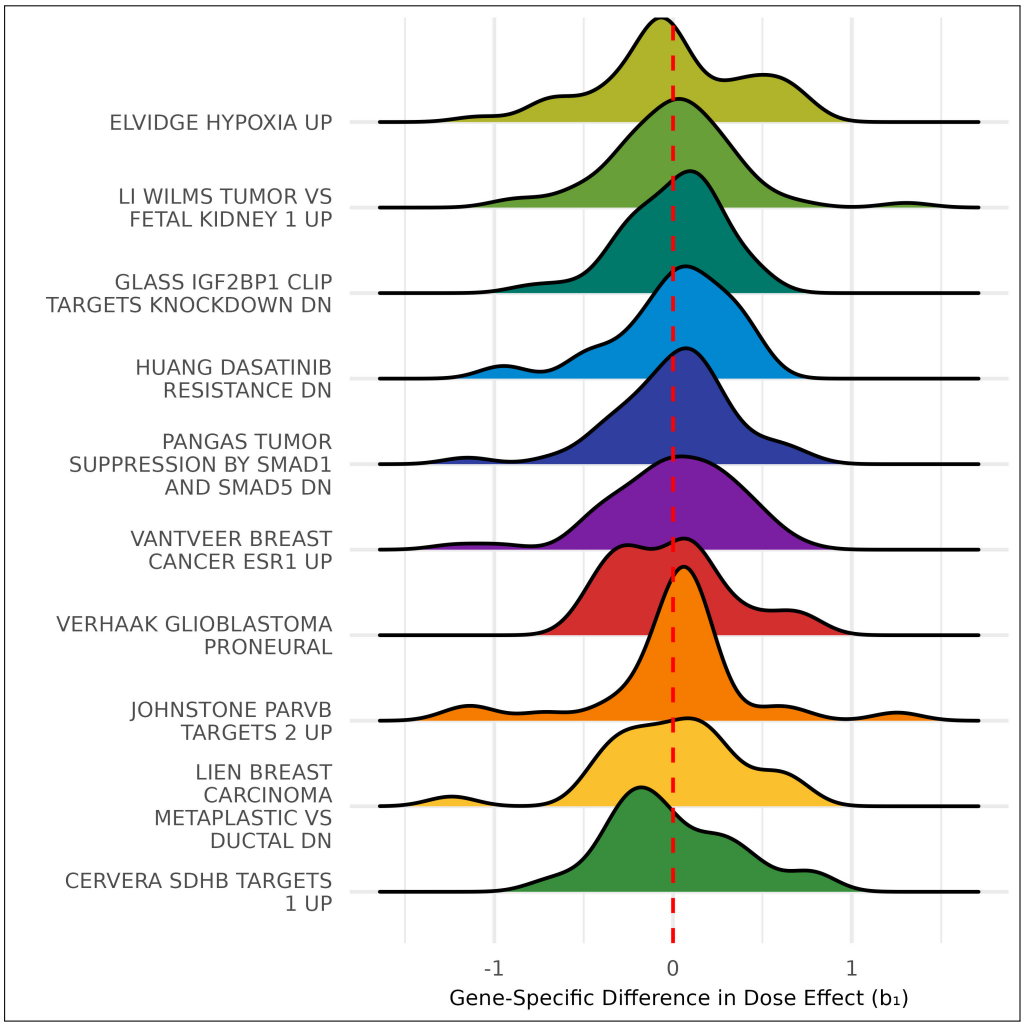

Figure 3: **Random Effect Distribution.** The distribution of gene random effects for each pathway can be visualized to assess dose-response, with wider distributions indicating greater variability in gene responses, and narrower distributions indicating more consistent gene trends within the pathways.

The distribution of random effects is visualized to help researchers assess the variability within each pathway (Fig. 3). Wider distributions indicate greater variability, while narrower distributions suggest more consistent gene responses across doses. This feature helps to identify pathways where a subset of genes shows distinct responses to chemical exposure, providing deeper insights into the biological processes affected by the dose.

2.5.2 Model Validation

After the models are fitted, **DoseRider** compares each model's performance to the null model using statistical metrics, including AICc and p-values. Only models that demonstrate both a significantly lower AICc and a meaningful p-value are retained for further analysis. This rigorous validation ensures that the dose-response relationships detected are both statistically significant and biologically meaningful.

To further enhance the accuracy of the BMD calculations, **DoseRider** employs bootstrapping methods, generating confidence intervals for BMD values by re-fitting the model to sub-sampled data. This process allows for robust estimation of both the upper and lower bounds of BMD values, offering researchers greater confidence in their results.

3 User Interface

The **DoseRider** interface is designed to be user-friendly and intuitive, guiding users through the entire process of setting up, analyzing, and visualizing dose-response relationships for multi-omics data. Below are the key sections and features of the interface, along with their functions.

3.1 Main Dashboard

The main dashboard is the starting point for users to access different features of **DoseRider**. It includes options to:

- **Explore pre-analyzed datasets**
- **Upload your own dataset**
- **View ongoing analyses**

3.1.1 Navigation Bar

The navigation bar provides quick access to various pages within **DoseRider**:

- **Home:** Returns the user to the dashboard.
- **Manual:** Links to the user manual, where detailed instructions can be found.
- **FAQ:** Access to frequently asked questions.
- **Datasets:** Directs users to the pre-analyzed datasets section.

3.2 Upload Your Own Data

The *Upload Your Own Data* page enables users to input their own experimental data for dose-response analysis. This page includes fields for:

- **Chemical name, time point, and omics type**
- **Annotation and metadata files** (required for analysis)
- **Dose/Concentration values** and units

Once the required fields are filled, the user can submit their dataset for analysis.

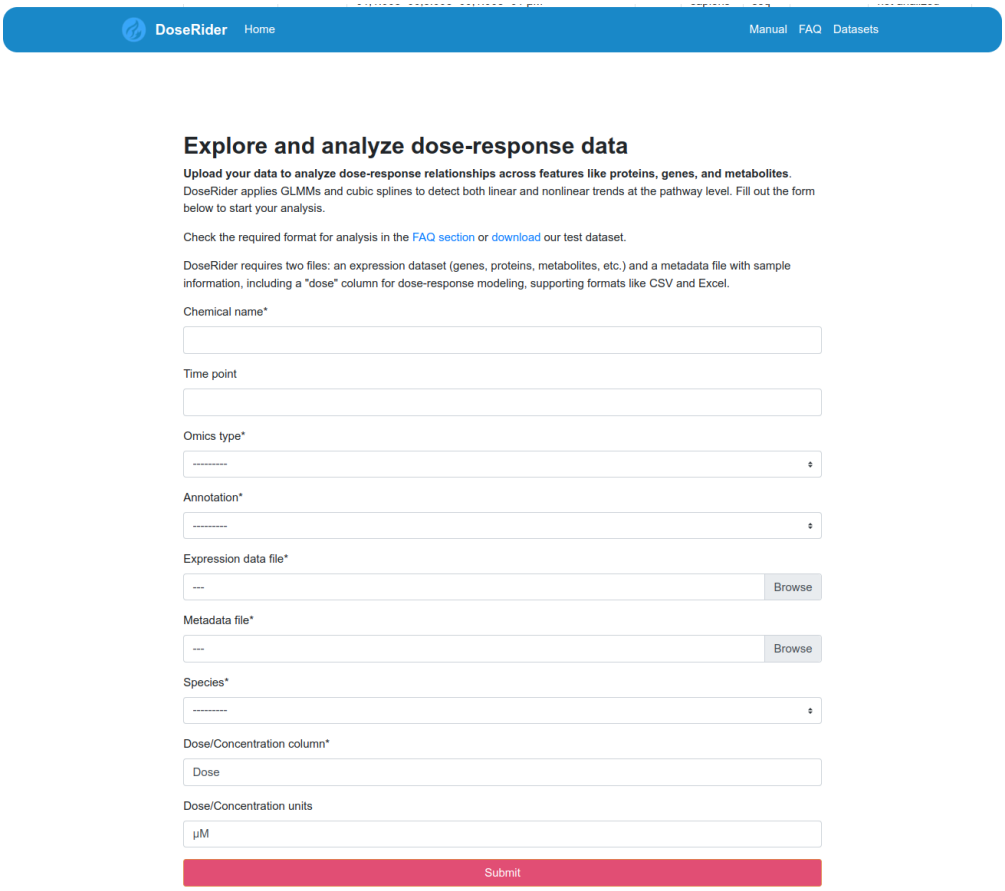

Figure 4: Upload your own data: Submit datasets, metadata files, and relevant dose information for analysis.

3.3 Explore Pre-Analyzed Multi-Omics Datasets

Alternative, the user can use the already pre-analized datasets. The *Explore Pre-Analyzed Multi-Omics Datasets* page provides access to pre-analyzed dose-response data for various omics types, species, and time points. Features include:

- Search and filter options to explore datasets by species, omics type, and dose range.
- Option to run new analyses or view results for previously analyzed datasets.

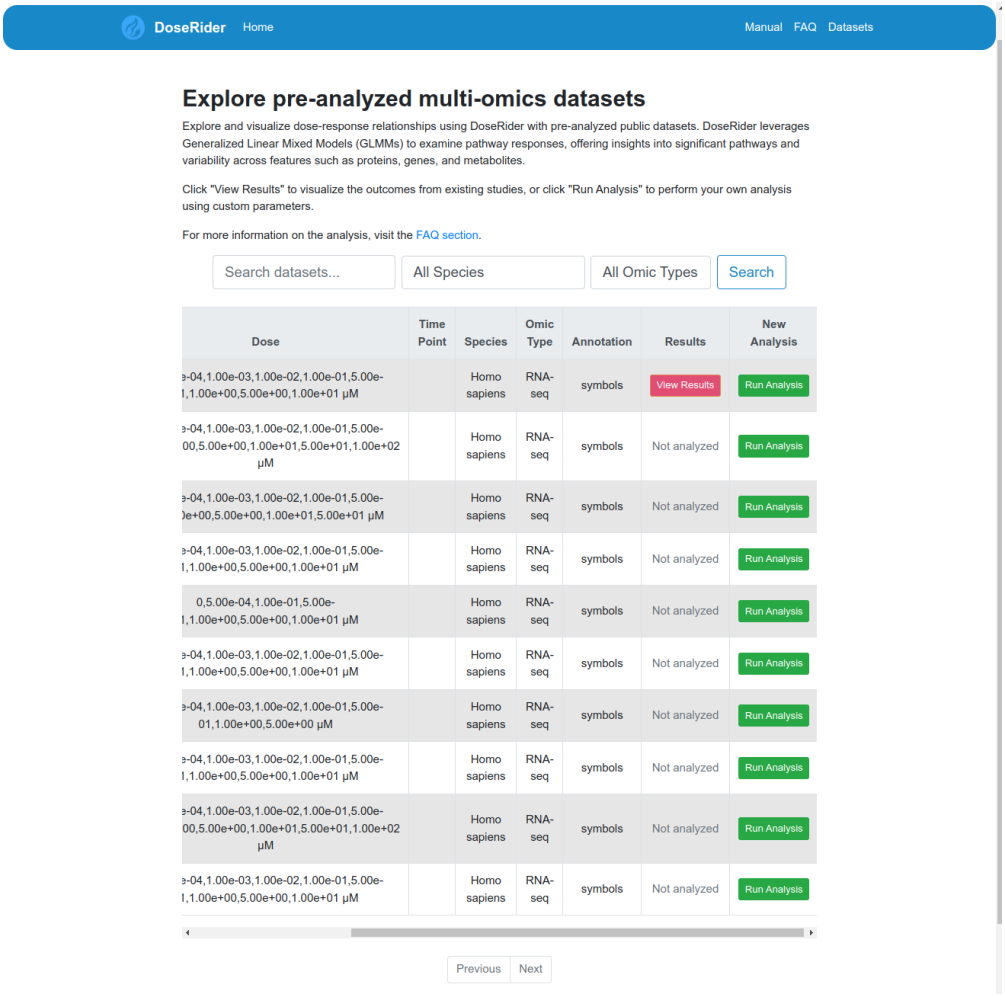

Figure 5: Pre-analyzed datasets page: Explore and visualize datasets based on various species, omics types, and experimental conditions.

3.4 Configure Your Dose-Response Analysis

The *Configure Your Dose-Response Analysis* page allows users to customize their analysis by adjusting feature sets, size filters, and transformation options. This section includes:

- **Dataset summary:** Information about the uploaded dataset, including gene set database and size filters.
- **Feature sets fields:** Allow to choose the databse set, such as, MSigDB, MIo, ConsensuPathDB, etc. And the set size.
- **Data Preprocessing:** Options to log-transform dose/concentration and intensity values, and apply feature filtering.
- **Model Fitting:** Select between *Linear*, *Non-linear (Fixed)*, or *Non-linear (Mixed)* models.
- **Viability Analysis:** Enable or disable viability analysis.

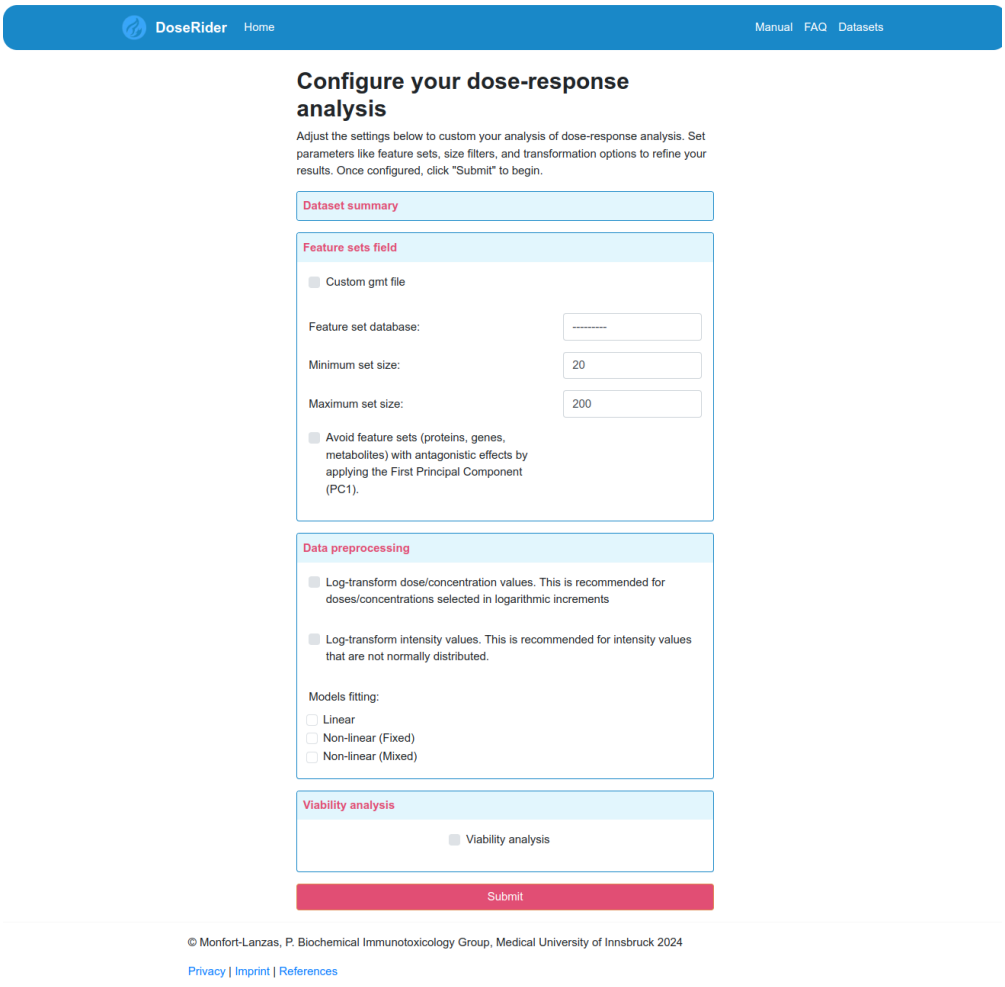

Figure 6: Configure your dose-response analysis: Users can adjust feature set sizes, data preprocessing options, and model fitting parameters.

3.5 Job Status Information

Once a job has been submitted for analysis, users can view the progress on the *Job Status Information* page. This page shows:

- Session URL for checking results later
- Job ID, status, description, and time elapsed

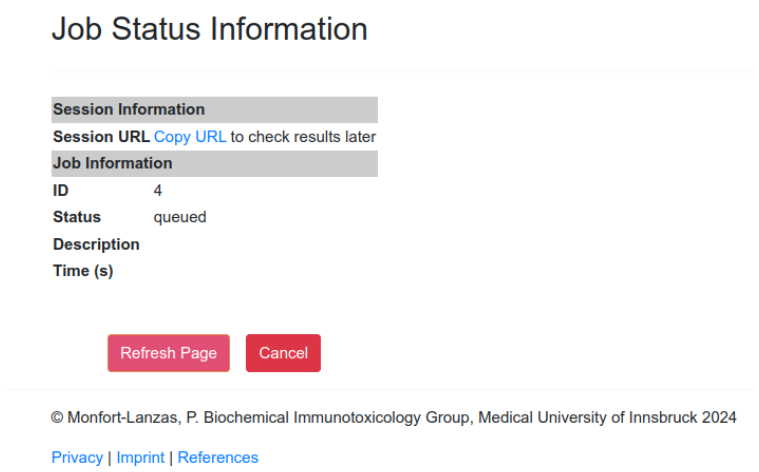

Figure 7: Job status information: Displays the current status of an analysis job, including job ID, elapsed time, and session link.

3.6 Geneset Dose-Response Model Results

The *Geneset Dose-Response Model Results* page presents detailed results of the dose-response models applied to the selected gene sets or pathways. This section allows users to evaluate the fit and performance of the different models (e.g., linear, non-linear fixed, non-linear mixed), along with important metrics like the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Benchmark Dose (BMD) values.

The results table on this page includes the following fields, providing a comprehensive overview of each gene set's dose-response model:

1. **Gene Set:** Name of the gene set or pathway analyzed (e.g., APOPTOSIS [HALLMARK]).
2. **Gene Set Size:** The total number of genes in the gene set.
3. **Genes in Gene Set:** The number of genes within the set that were included in the analysis.
4. **Best Model:** The model that best fits the data, selected from linear, non-linear fixed, or non-linear mixed models.
5. **Best Model P-value:** The p-value associated with the best-fitting model, indicating the statistical significance of the model.
6. **Best Model Adjusted P-value:** The adjusted p-value (e.g., using FDR) for the best model, correcting for multiple comparisons.
7. **Best Model AICc:** The corrected Akaike Information Criterion for the best model, balancing model fit with complexity.
8. **Best Model BIC:** The Bayesian Information Criterion for the best model, providing a second criterion for evaluating model fit.
9. **BMD Lower Bound:** The lower confidence interval for the Benchmark Dose, which represents the lowest dose with a significant response.
10. **BMD Median:** The median Benchmark Dose, indicating the dose where a benchmark level of response occurs.
11. **BMD Upper Bound:** The upper confidence interval for the Benchmark Dose, providing the upper range of significant response.
12. **BMD Mean:** The mean Benchmark Dose, offering an average estimate across the data.
13. **Null AIC:** The Akaike Information Criterion for the null model (no dose-response relationship).
14. **Null AICc:** The corrected Akaike Information Criterion for the null model.
15. **Null BIC:** The Bayesian Information Criterion for the null model.
16. **Null df:** The degrees of freedom for the null model.
17. **Linear AIC:** The Akaike Information Criterion for the linear model.
18. **Linear AICc:** The corrected Akaike Information Criterion for the linear model.
19. **Linear BIC:** The Bayesian Information Criterion for the linear model.
20. **Linear df:** The degrees of freedom for the linear model.
21. **Linear ICC:** The intraclass correlation coefficient for the linear model, assessing the consistency of gene responses within the pathway.
22. **Non-linear Fixed AIC:** The Akaike Information Criterion for the non-linear fixed effects model.
23. **Non-linear Fixed AICc:** The corrected Akaike Information Criterion for the non-linear fixed effects model.

24. **Non-linear Fixed BIC:** The Bayesian Information Criterion for the non-linear fixed effects model.
25. **Non-linear Fixed df:** The degrees of freedom for the non-linear fixed effects model.
26. **Non-linear Fixed ICC:** The intraclass correlation coefficient for the non-linear fixed effects model.
27. **Non-linear Mixed AIC:** The Akaike Information Criterion for the non-linear mixed effects model.
28. **Non-linear Mixed AICc:** The corrected Akaike Information Criterion for the non-linear mixed effects model.
29. **Non-linear Mixed BIC:** The Bayesian Information Criterion for the non-linear mixed effects model.
30. **Non-linear Mixed df:** The degrees of freedom for the non-linear mixed effects model.
31. **Non-linear Mixed ICC:** The intraclass correlation coefficient for the non-linear mixed effects model.
32. **P-value (Linear):** The p-value for the linear model, indicating the significance of the dose-response relationship.
33. **P-value (Non-linear Fixed):** The p-value for the non-linear fixed effects model.
34. **P-value (Non-linear Mixed):** The p-value for the non-linear mixed effects model.
35. **Adjusted P-value (Linear):** The adjusted p-value for the linear model, correcting for multiple hypothesis testing.
36. **Adjusted P-value (Non-linear Fixed):** The adjusted p-value for the non-linear fixed effects model.
37. **Adjusted P-value (Non-linear Mixed):** The adjusted p-value for the non-linear mixed effects model.
38. **Optimal Clusters:** The number of optimal gene clusters determined for the pathway, based on the dose-response behavior of the genes.

Users can filter results based on adjusted p-values or model types to focus on significant dose-response relationships. The results table can be exported in various formats, including CSV, Excel, and TSV, for further analysis.

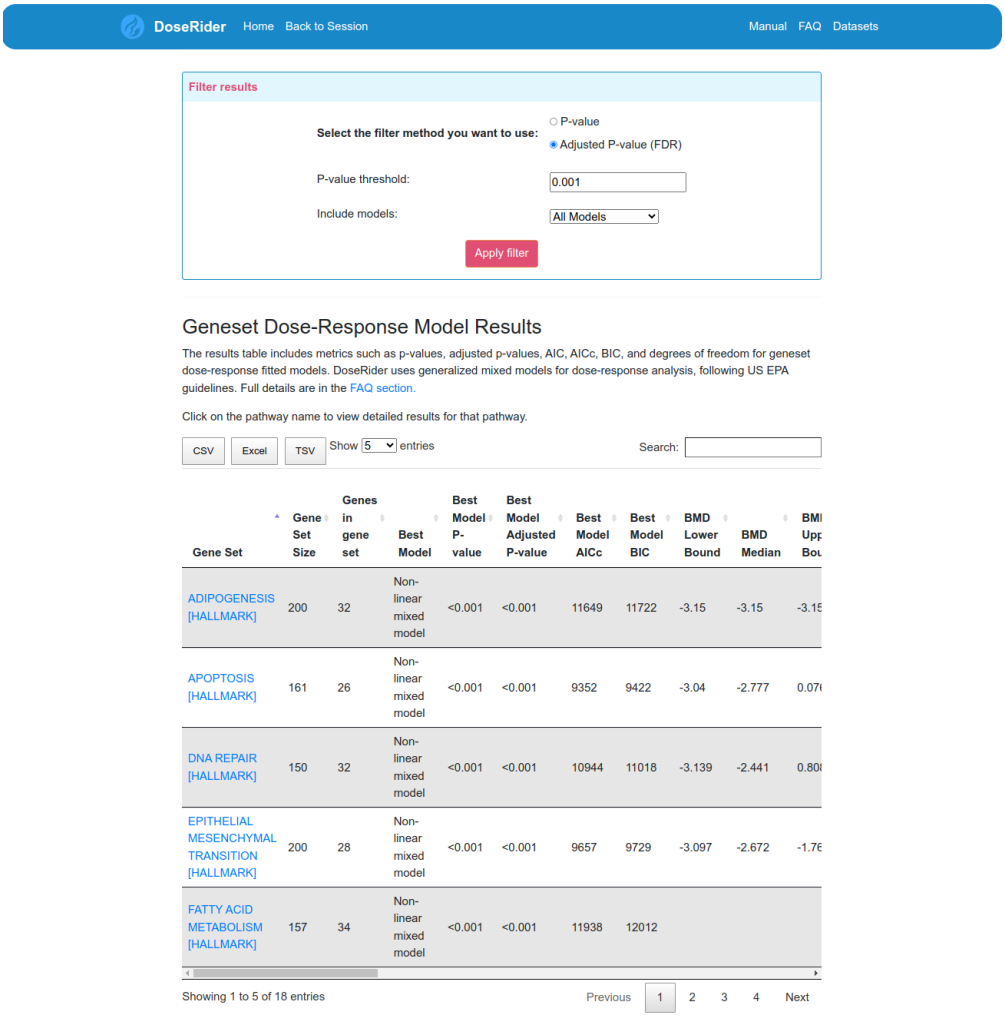

Figure 8: Geneset dose-response model results: Summary of the best models, including statistical metrics and BMD values.

3.7 Visualization Categories

DoseRider offers a comprehensive set of visualizations to interpret dose-response behavior at both the dataset and pathway levels. Visual outputs are grouped into two categories:

Dataset-Level Visualizations

These figures summarize global trends and modeling results across all pathways:

1. **Dotplot of Top Pathways:** Ranks pathways by -$\log_{10}$(adjusted p-value). Larger points represent higher significance, helping users quickly identify responsive pathways.
2. **Dose Response Heatmap:** Displays mean gene expression responses across doses for significant pathways. Color intensity indicates expression magnitude.
3. **Gene Set Random Effects:** Highlights gene-level variability within pathways by visualizing random effect estimates from mixed models.
4. **Top Pathway Responses:** Shows predicted dose-response curves for the most significant pathways based on best-fit models, reflecting overall trend direction and effect magnitude.
5. **BMD Confidence Intervals:** Displays 95% bootstrap confidence intervals for Benchmark Dose (BMD) estimates, indicating dose levels with statistically significant changes.
6. **BMD Density and Peaks:** Illustrates the density of BMD values across pathways to identify dose regions with high biological activity.
7. **TCD Density and Peaks:** Shows the distribution of Trend Change Doses (TCDs), which mark points of curvature change (maxima/minima) in the dose-response trajectory, based on spline derivatives.
8. **TCD1 Confidence Intervals:** Provides 95% bootstrap intervals for TCD estimates, capturing uncertainty around dose thresholds where expression direction shifts.

Users can explore these plots interactively using tabbed views in the interface.

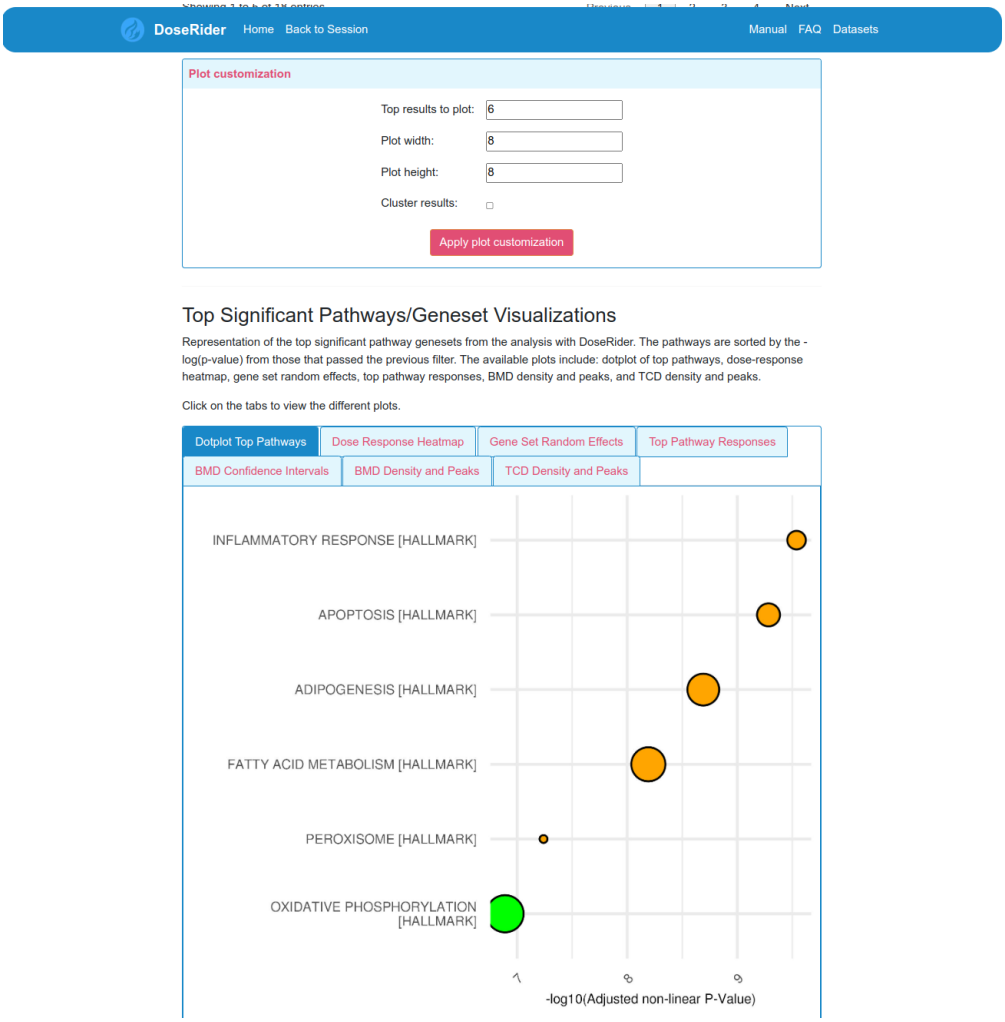

Figure 9: Dotplot of top pathways ranked by statistical significance.

3.7.1 Pathway-Level Visualizations

These figures provide a detailed view of individual gene set behavior:

- **Gene-Specific Response and Variability:** Displays gene-level deviations via random effects from the mixed model.
- **Dose Response Heatmap (per Pathway):** Shows gene-wise expression across doses for a selected pathway.
- **Pathway Responses:** Visualizes the modeled dose-response curve for a given pathway, integrating all genes.

Users can adjust plot size, clustering, and other settings to tailor these visualizations.

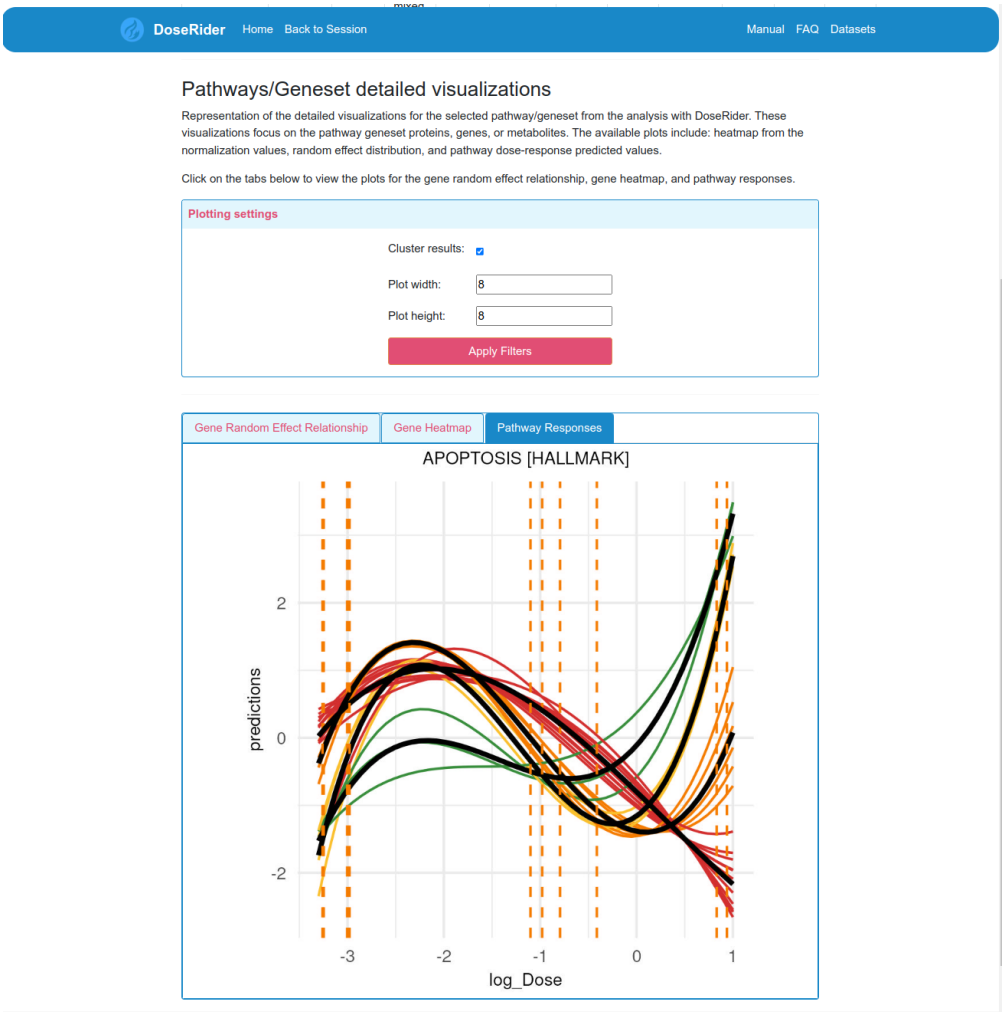

Figure 10: Pathway-level visualizations: random effects, gene heatmaps, and smoothed responses.

4 Troubleshooting

The DoseRider platform is designed to be intuitive and user-friendly. However, issues may occasionally arise during data upload, analysis, or visualization. Below are some common troubleshooting tips for resolving such issues.

4.1 Common Issues and Solutions

- **Problem: Dataset upload fails.**
  - **Solution:** Ensure that your dataset is in the correct format (CSV, TSV, or Excel). Check that the columns are properly labeled, and the dose information is provided in a separate column. If using metadata, ensure that it corresponds accurately with the sample identifiers in the expression data.
- **Problem: Analysis job is stuck in the "queued" status.**
  - **Solution:** This may occur if the server is experiencing high demand. Refresh the page after a few minutes. If the issue persists, cancel the job and resubmit the analysis. Ensure your internet connection is stable during submission.
- **Problem: No results are returned after running the analysis.**
  - **Solution:** Verify that the correct columns for dose and sample identifiers are properly labeled in the metadata file. Double-check the gene set size thresholds and ensure that the dataset contains sufficient variability across doses.
- **Problem: Visualizations do not display or are incomplete.**
  - **Solution:** Ensure that JavaScript is enabled in your browser, and try clearing the browser cache. If the problem persists, try running the analysis in a different browser or updating your current browser to the latest version.
- **Problem: Benchmark Dose (BMD) or Trend Change Dose (TCD) values appear inconsistent or incorrect.**
  - **Solution:** Check the dose units in your metadata file (e.g., µM or mg/L) and ensure they match across all samples. Make sure the data has been preprocessed correctly (e.g., log-transformation of doses if applicable) to avoid skewed results.
- **Problem: Unable to download results in CSV/Excel/TSV format.**
  - **Solution:** Ensure pop-ups are not blocked in your browser settings, as some browsers may prevent automatic downloads. If the problem persists, try exporting the results using a different format.
- **Problem: Unexpected error messages during analysis.**
  - **Solution:** These could be caused by incompatible file formats or errors in the dataset. Double-check the formatting of the uploaded data and ensure the metadata aligns with the expression data. Refer to the file format guidelines in the user manual. Contact technical support if the issue persists.

4.2 Contact Support

If these troubleshooting steps do not resolve your issue, you can reach out to the **DoseRider** technical support team at *pablo.monfort@i-med.ac.at*. Please include details about the issue, such as the dataset format, error messages received, and the steps leading up to the issue.

5 Acronyms

Akaike Information Criterion

Corrected Akaike Information Criterion

Bayesian Information Criterion

Benchmark Dose

Comma-Separated Values

False Discovery Rate

Generalized Linear Mixed Models

Intraclass Correlation Coefficient

JavaScript Object Notation

Principal Component 1

RNA Sequencing

Summarized Experiment Matrix

Single Nucleotide Polymorphism

Tab-Separated Values

Trend Change Dose