

No One is Left Behind: Bridging the Proteomics Divide with Mass Dynamics 2.0's Dataset Service

Anna Quagliari¹, Aaron Triantafyllidis¹, Bradley Green¹, Mark R. Condina^{1,2}, Paula Burton Ngov¹, Giuseppe Infusini¹ and Andrew I. Webb^{1,3,4}

¹Mass Dynamics, Melbourne, Victoria 3000, Australia, ²Clinical & Health Sciences, University of South Australia, Adelaide, SA 5095, Australia, ³The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia, ⁴Department of Medical Biology, University of Melbourne, Melbourne, Victoria 3010, Australia

Introduction

Mass Dynamics 2.0 (MD 2.0) exists to lower the computational barrier leading to meaningful insight generation and to efficient collaboration on proteomics datasets of any scale¹.

Here, we introduce the new Dataset Service in MD 2.0, meticulously engineered to cater to these rapidly evolving demands so as to ensure that advancements in the fields are accessible to all researchers. The Dataset service allows the implementation of specific proteomics workflows, such as Dose-Response analysis.

MD 2.0's Dataset Service

The Dataset Service allows processing and re-processing of various types of analyses. Its conception was pushed by careful thinking of the data science approach needed in Bioinformatics, allowing:

- Researchers to compare different types of analyses at every stage of the process: QC, normalization, imputation, group comparisons etc..
- To easily scale the number of analyses provided as new ones are requested
- A data pipeline that supports the storage and retrieval of any dataset analysed for easy comparing using the flexible modular interface previously introduced
- A user-friendly interface to trigger multiple analyses simultaneously
- The ability to add several layers of samples metadata

What does the Dataset Service look like?

MD 2.0's existing modular workspace for flexible visualizations (Figure 1) is now coupled with a flexible and extensible engine to create, store and access multiple datasets (Figure 2):

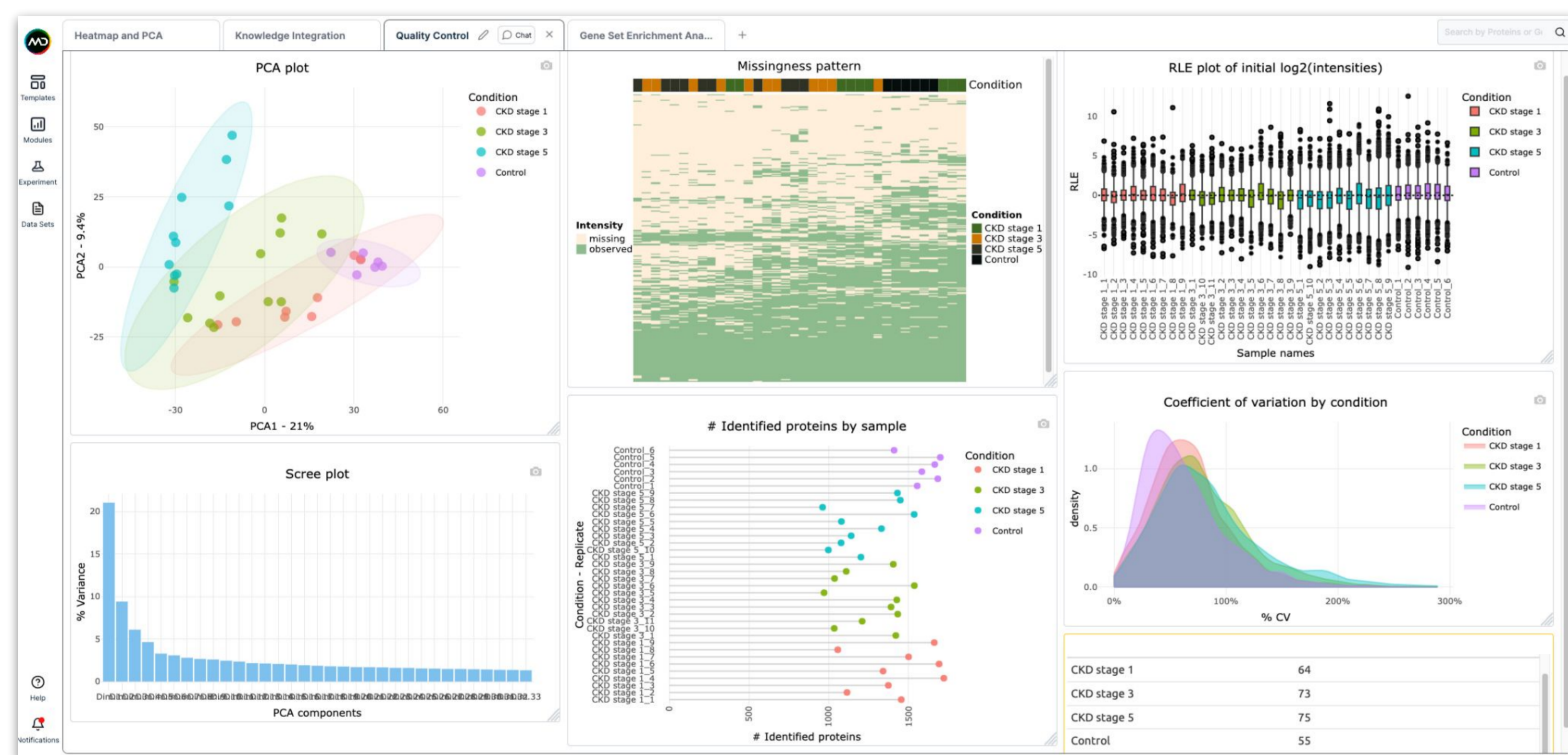


Figure 1: Example output from MD2.0 after uploading a dataset and building a template consisting of various visualizations.

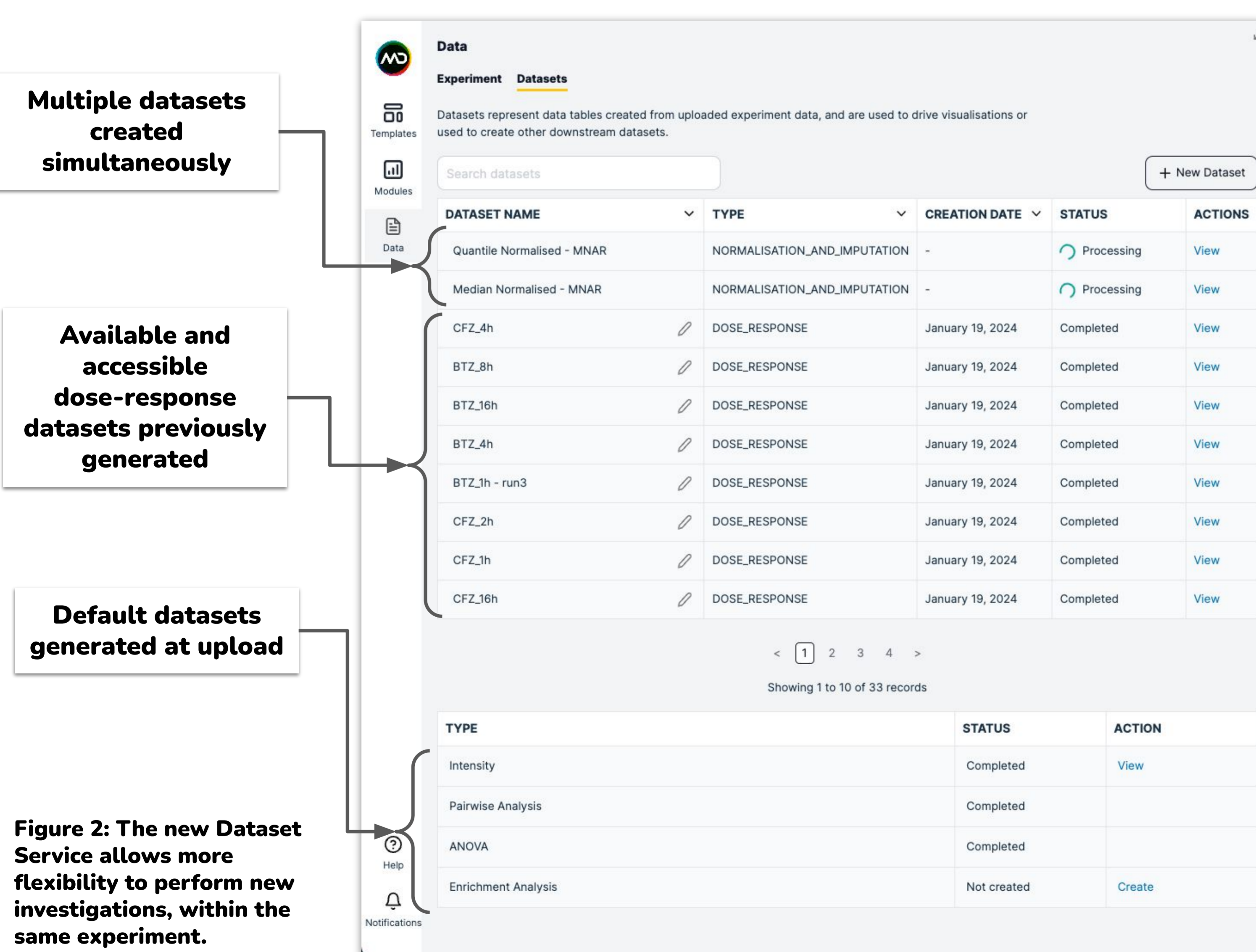


Figure 2: The new Dataset Service allows more flexibility to perform new investigations, within the same experiment.

Case Study: Streamline dose-response investigations on multiple compounds

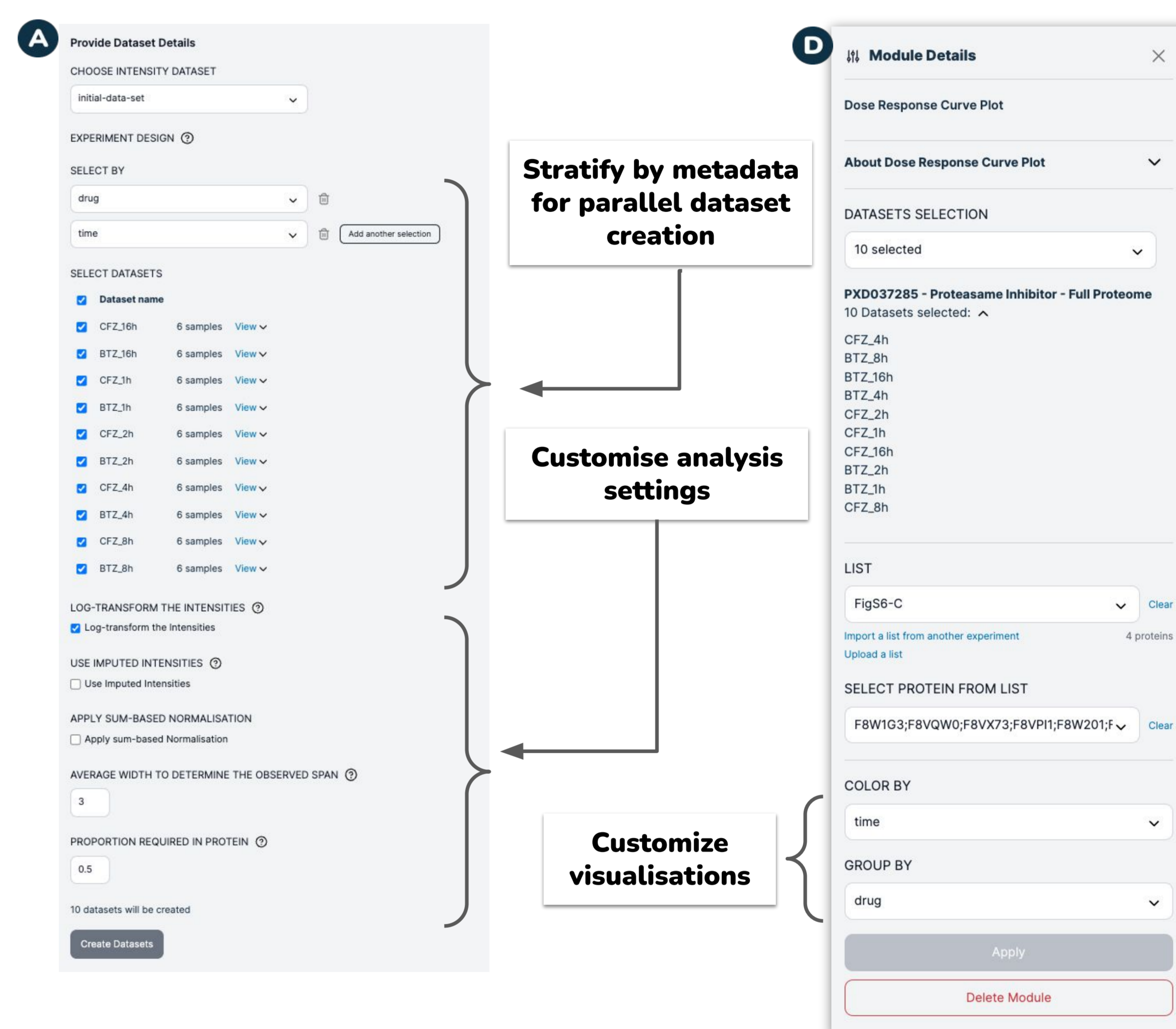
Data: MaxQuant Proteasome dose-response data (PXD037285²) was uploaded to MD 2.0.

Dose-Response Design: Multiple Myeloma cell line RPMI 8226 was treated with several doses of two drugs, bortezomib (BTZ) and carfilzomib (CFZ). The dose-response data was collected at 5 different time points.

Methods: A dose-response dataset consists in fitting regression models between the intensity of each protein and non-negative doses of an exposure variable interest. Models are fitted using the 4-parameter log-logistic regression in the drc R package³ and visualisations are generated using Plotly⁴ and Matplotlib⁵.

Easily generate and investigate 10 dose-response datasets in MD 2.0

1. Upload sample metadata to stratify datasets creation & curve visualizations
2. Run all dose-response analyses simultaneously (A)



3. Visualize & interrogate the results across datasets (B-D)

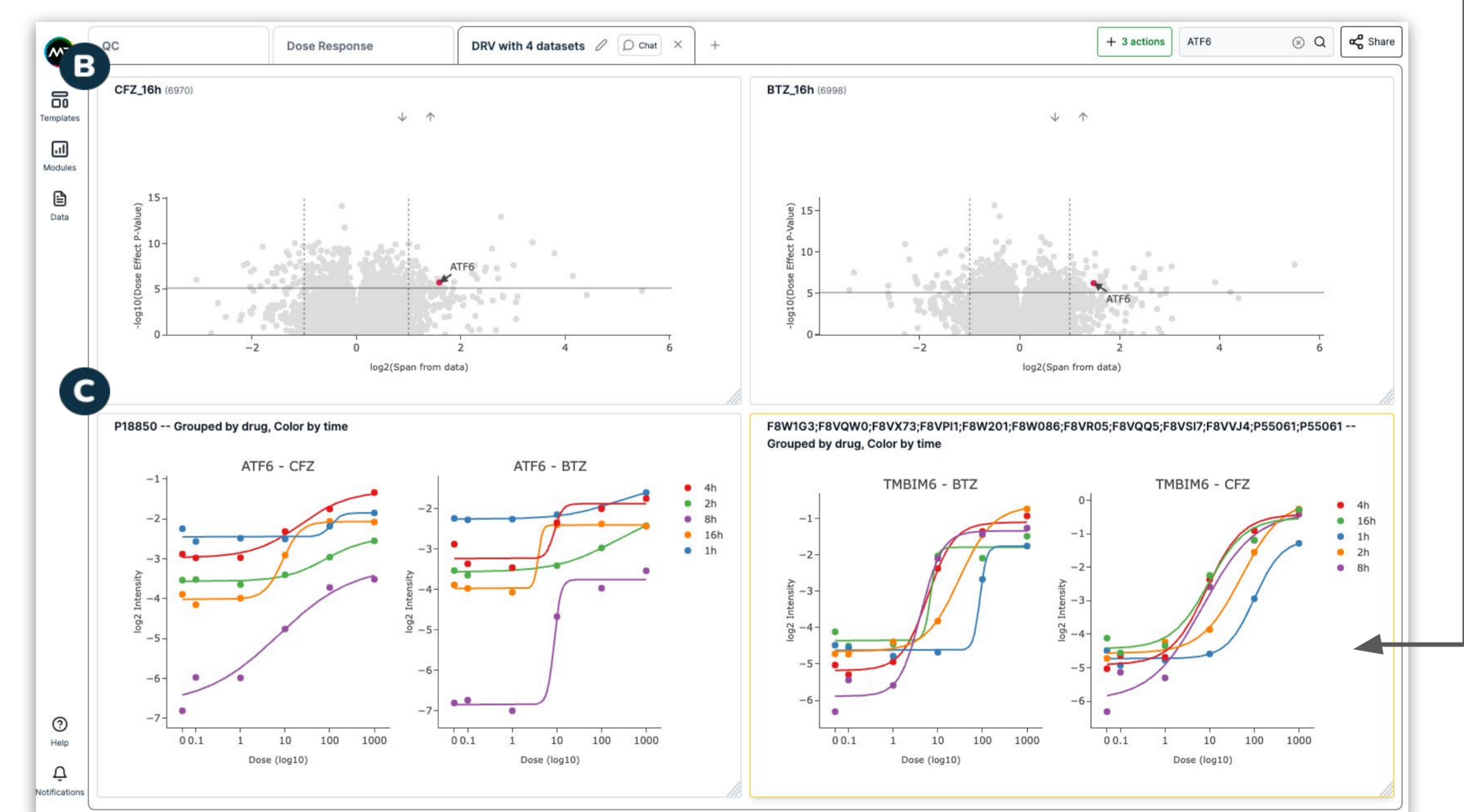


Figure 3: Dose-response analysis workflow with the Dataset Service in MD 2.0.

Future Directions

- Broaden statistical analyses options for dataset service, e.g. time series, imputations
- More flexibility with customized templates and ability to integrate new templates and analysis with community-based input;
- Workflows to support post-translational modifications (PTMs), including phosphorylation.

Request a Free Trial

Keen to take Mass Dynamics for a free test drive? Simply click on the QR code and you could be exploring our discovery service for up to 14 days.



Figures

Figure 1: Example output from MD2.0 after uploading a dataset and building a template consisting of various visualizations. The user has defined a summary of visuals to assess data quality which include: (A) Principal Components Analysis (PCA); (B) Scree plots of PCA; (C) Missingness heatmap; (D) Relative Log Expression (RLE) plots; (E) Number of identified proteins; (F) CV distribution plot coloured by condition; (G) CV distribution table by condition.

Figure 2: The new Dataset Service allows more flexibility to perform new investigations, within the same experiment. (A) The dataset module can be accessed by selecting the "Data" icon in the MD 2.0 interface. Any new Datasets created are listed. (B) Specific analyses are displayed as separate datasets, and are dependent on the experiment design.

Figure 3: Dose-response analysis workflow with the Dataset Service in MD 2.0. (A) Setup and create multiple dose-response analysis using sample metadata and define settings of the analysis; (B) Any new Datasets created are listed. (B) Visualize dose-response results for all proteins in the experiments across multiple datasets using dose-response volcano plots; (C) Visualize dose-response curves stratified by dataset metadata for target proteins; (D) Module settings of dose-response curve visualization.

References

1. Quagliari A, et al. Mass Dynamics 2.0: An improved modular web-based platform for accelerated proteomics insight generation and decision making. *BioRxiv* (2022) 2. Zecha J, et al. Deciphering drug actions and protein modifications by dose- and time-resolved proteomics. *Science* (2023)
3. Ritz C, et al. Dose-response analysis using R. *Plos One* (2015) 4. Inc, P. T. Collaborative data science. Montreal: Plotly Technologies Inc Montreal (2015) 5. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* (2007)