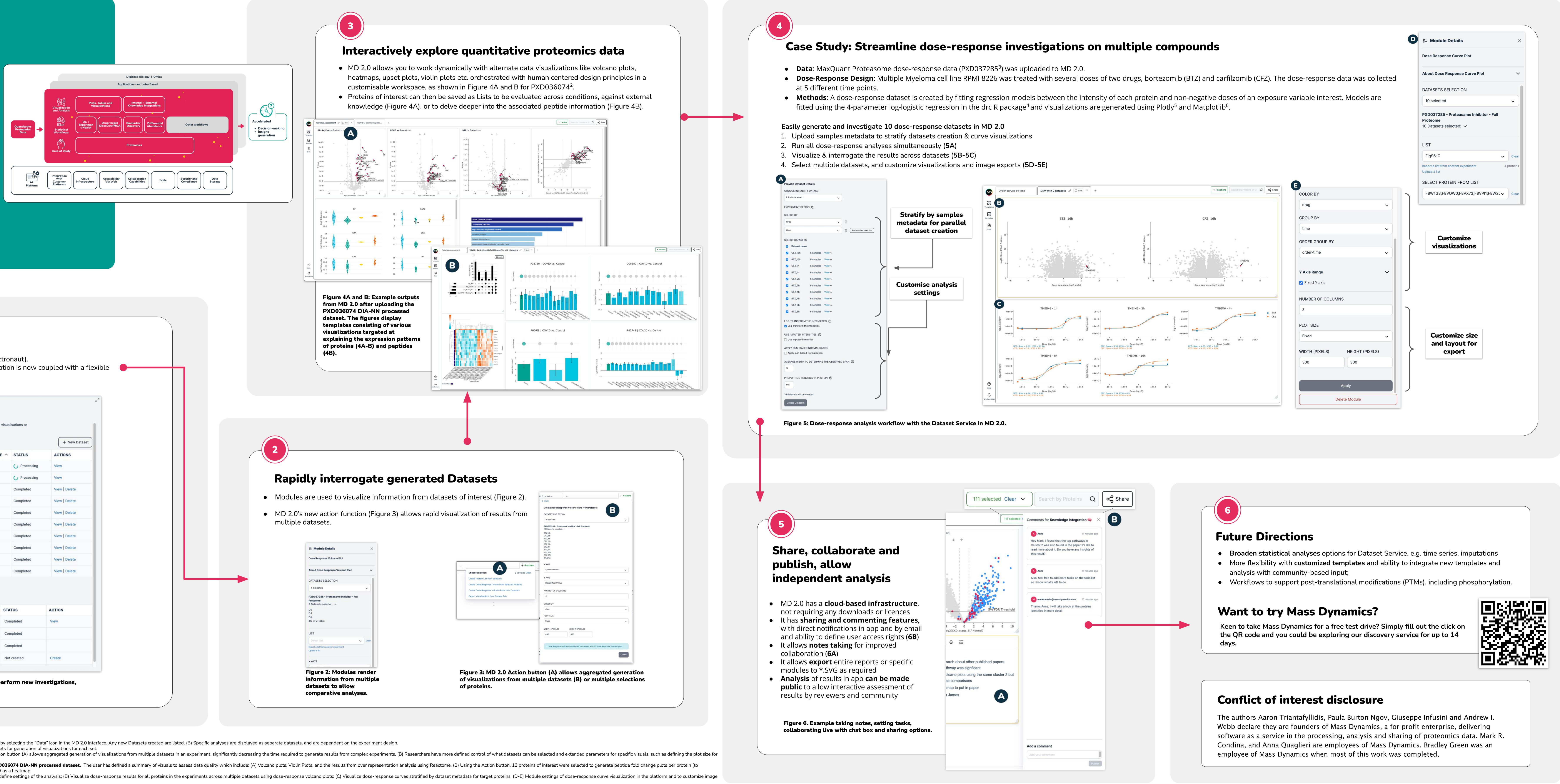# Integrate, Automate and Interrogate Proteomics Workflows with MD 2.0 Dataset Service

Anna Quaglieri[1], Aaron Triantafyllidis[1], Bradley Green[1], Mark R. Condina[1,2], Paula Burton Ngov[1], Giuseppe Infusini[1] and Andrew I. Webb[1,3,4]

[1]Mass Dynamics, Melbourne, Victoria 3000, Australia, [2]Clinical & Health Sciences, University of South Australia, Adelaide, SA 5095, Australia, [3]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia, [4]Department 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 efficient collaboration on proteomics datasets of any scale[1].

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

Case studies: How MD 2.0 facilitates multi-experiment functionality

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 analyzed** 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**

1. What does the Dataset Service look like?

- Upload data from pre-processed analysis (e.g. MaxQuant, Bruker ProteoScape™, DIA-NN, Spectronaut).
- MD 2.0's existing capability to combine modules or use templated analyses for flexible visualization is now coupled with a flexible and extensible engine to create, store and access multiple datasets (Figure 1).

Figure 1: The new Dataset Service allows more flexibility to perform new investigations, within the same experiment.

2. Rapidly interrogate generated Datasets

- Modules are used to visualize information from datasets of interest (Figure 2).
- MD 2.0's new action function (Figure 3) allows rapid visualization of results from multiple datasets.

Figure 2: Modules render information from multiple datasets to allow comparative analyses.

Figure 3: MD 2.0 Action button (A) allows aggregated generation of visualizations from multiple datasets (B) or multiple selections of proteins.

3. Interactively explore quantitative proteomics data

- MD 2.0 allows you to work dynamically with alternate data visualizations like volcano plots, heatmaps, upset plots, violin plots etc. orchestrated with human centered design principles in a customisable workspace, as shown in Figure 4A and B for PXD036074[2].
- Proteins of interest can then be saved as Lists to be evaluated across conditions, against external knowledge (Figure 4A), or to delve deeper into the associated peptide information (Figure 4B).

Figure 4A and B: Example outputs from MD 2.0 after uploading the PXD036074 DIA-NN processed dataset. The figures display templates consisting of various visualizations targeted at explaining the expression patterns of proteins (4A-B) and peptides (4B).

4. Case Study: Streamline dose-response investigations on multiple compounds

- **Data**: MaxQuant Proteasome dose-response data (PXD037285[3]) 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 is created by 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 package[4] and visualizations are generated using Plotly[5] and Matplotlib[5].

Easily generate and investigate 10 dose-response datasets in MD 2.0
1. Upload samples metadata to stratify datasets creation & curve visualizations
2. Run all dose-response analyses simultaneously (5A)
3. Visualize & interrogate the results across datasets (5B-5C)
4. Select multiple datasets, and customize visualizations and image exports (5D-5E)

Figure 5: Dose-response analysis workflow with the Dataset Service in MD 2.0.

5. Share, collaborate and publish, allow independent analysis

- MD 2.0 has a **cloud-based infrastructure**, not requiring any downloads or licences
- It has **sharing and commenting features**, with direct notifications in app and by email and ability to define user access rights (6B)
- It allows **notes taking** for improved collaboration (6A)
- It allows **export** entire reports or specific modules to *.SVG as required
- **Analysis** of results in app **can be made public** to allow interactive assessment of results by reviewers and community

Figure 6: Example taking notes, setting tasks, collaborating live with chat box and sharing options.

6. 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.

Want to try Mass Dynamics?

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

Conflict of interest disclosure

The authors Aaron Triantafyllidis, Paula Burton Ngov, Giuseppe Infusini and Andrew I. Webb declare they are founders of Mass Dynamics, a for-profit enterprise, delivering software as a service in the processing, analysis and sharing of proteomics data. Mark R. Condina, and Anna Quaglieri are employees of Mass Dynamics. Bradley Green was an employee of Mass Dynamics when most of this work was completed.

Figures
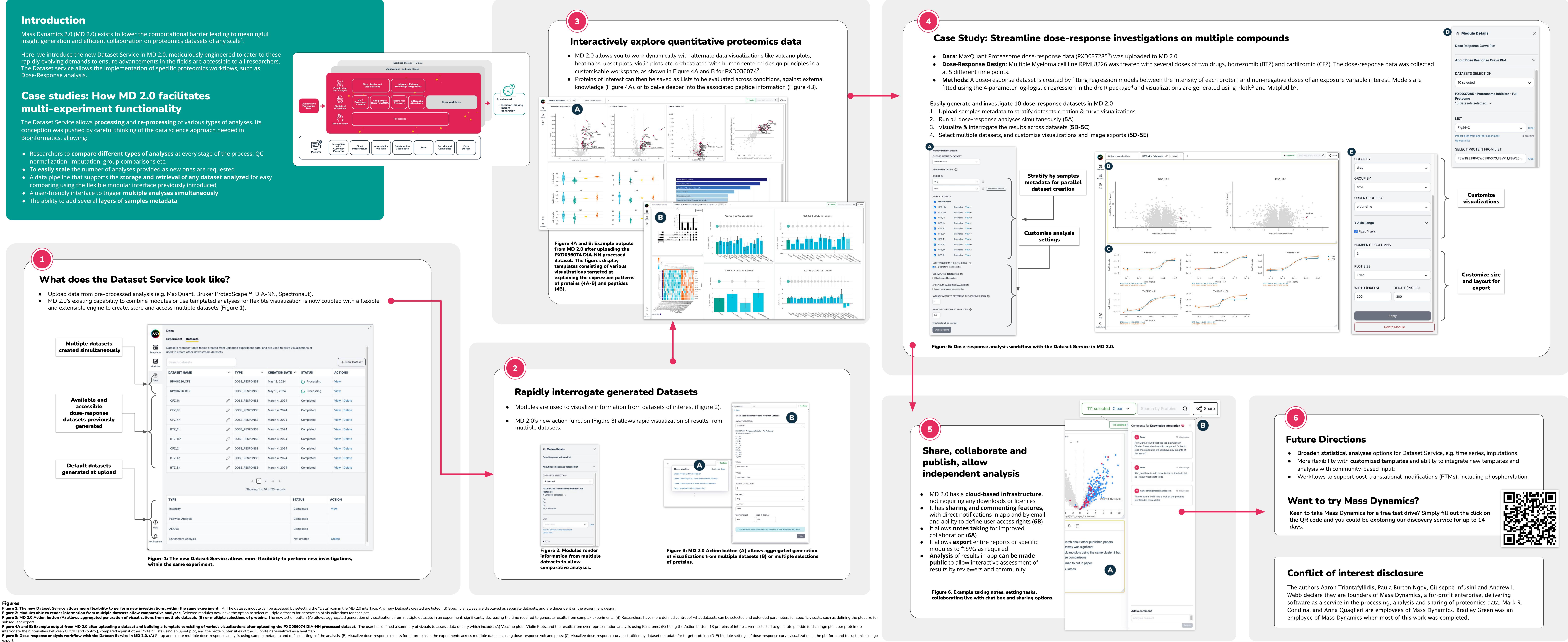Figure 1: The new Dataset Service allows more flexibility to perform new investigations, within the same experiment. (A) The dataset module can be accessed by selecting the "Data" icon in the MD 2.0 interface. Any new Datasets created are listed. (B) Specific analyses are displayed as separate datasets, and are dependent on the experiment design.
Figure 2: Modules able to render information from multiple datasets allow comparative analyses. Selected modules now have the option to select multiple datasets for generation of visualizations for each set.
Figure 3: MD 2.0 Action button (A) allows aggregated generation of visualizations from multiple datasets (B) or multiple selections of proteins. The new action button (A) allows aggregated generation of visualizations from multiple datasets in an experiment, significantly decreasing the time required to generate results from complex experiments. (B) Researchers have more defined control of what datasets can be selected and extended parameters for specific visuals, such as defining the plot size for subsequent export.
Figure 4A and B: Example output from MD 2.0 after uploading a dataset and building a template consisting of various visualizations after uploading the PXD036074 DIA-NN processed dataset. The user has defined a summary of vizuals to assess data quality which include: (A) Volcano plots, Violin Plots, and the results from over representation analysis using Reactome. (B) Using the Action button, 13 proteins of interest were selected to generate peptide fold change plots per protein (to interrogate their intensities between COVID and control), compared against other Protein Lists using an upset plot, and the protein intensities of the 13 proteins visualized as a heatmap.
Figure 5: 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) 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-E) Module settings of dose-response curve visualization in the platform and to customize image export.
Figure 6: Example taking notes, setting tasks, collaborating live with chat box and sharing experiment on MD 2.0. (A) Chat to for live collaboration with collaborators that have access to the experiment; (B) "Share" button to share experiment with collaborators (covered by the live chat in main panel).

References
1. Quaglieri A, et al. Mass Dynamics 2.0: An improved modular web-based platform for accelerated proteomics insight generation and decision making. BioRxiv (2022) 2. Wang Z, et al. The human host response to monkeypox infection: a proteomic case study. EMBO Mol Med. 2022 Nov 8;14(11):e16643. 3. Zecha J, et al. Decrypting drug actions and protein modifications by dose- and time-resolved proteomics. Science (2023) 4. Ritz C, et al. Dose-response analysis using R. Plos One (2015) 5. Inc, P. T. Collaborative data science. Montreal: Plotly Technologies Inc Montreal (2015) 6. Hunter, J. D. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. (2007)