Empower your proteomics research: streamline integration, automation, and analysis with MD 2.0 Dataset Service

Authors

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

View Session Details

Institutes

[1]Mass Dynamics, Melbourne, Australia; [2]University of South Australia, Adelaide, Australia; [3]WEHI, Parkville, Australia; [4]University of Melbourne, Parkville, Australia

Introduction

The proliferation of complex proteomics experiments and the advent of plenty of new workflows often developed by and for computational biologists act as a barrier to efficiently generating meaningful biological results, particularly for those not trained in computational methods. Mass Dynamics 2.0 (MD 2.0) exists to make current and new methodologies accessible to all researchers in an intuitive web and cloud-based collaborative environment. We introduce here the new Dataset Service in MD 2.0, engineered to incorporate and compare new workflows easily and to be adopted by researchers with any computational skill level thanks to its easy-to-use user interface.

Methods

In MD 2.0, a dataset encapsulates any computational result, such as protein intensities or enrichment analysis results. The Dataset Service facilitates analysis and re-analysis through an intuitive user interface and allows seamless integration of any new scientific workflow. Every iteration and detail of the data is securely stored for easy evaluation and comparison in MD 2.0's interactive and modular visualization workspace. The service now includes features like dose-response analyses (implemented from the drc R package, Ritz et al 2015), multiple normalization and imputation methods, and dimensionality reduction techniques.

Preliminary Data

This study illustrates the MD 2.0 Dataset Service's capabilities in expediting the creation of multiple dose-response analyses. Dose-response analyses are widely adopted in the drug discovery pipeline to screen the effect that several doses of compounds of interest have on protein intensities. The ability to automate and standardize the creation of several dose-response analyses and cohesively interrogate the results has the potential to streamline and speed up the process of target compounds selection. Using the proteasome inhibitor dataset from Zecha J *et al* 2023, we demonstrate how a typical iterative approach for quality control and intensity normalization is achieved in MD 2.0 and the automated generation of ten dose-response datasets. The proteasome dataset has protein intensities from the Multiple Myeloma cell line RPMI 8226 treated with several doses of two drugs, bortezomib and carfilzomib, and collected at five different time points. After upload and quality control, the distributions of the protein intensities are assessed using interactive boxplots and PCA before and after applying different normalization and imputation methods. Subsequently, the creation of the ten dose-response datasets is triggered simultaneously using the new Dataset Service user interface that automates

sample selection for analyses using the available metadata i.e. time, compound. Once the dose-response datasets are generated, the results are visualized in several dose-response volcano plots summarizing using various statistics such as the span, DMAX, EC50, and P-value of the effect. Proteins significantly affected by the treatment are prioritized via selection on volcano plots and simultaneously visualized in dose-response curves across all ten datasets.

Novel Aspect

MD 2.0 Dataset Service unlocks proteomics data analysis to any researcher allowing interrogation of analysis iterations in an interactive environment.

Conflict of Interest Disclosure

The authors declare no competing financial interest.