## Exploring protein expression across multiple experiments using Mass Dynamics



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

<sup>1</sup>Mass Dynamics, Melbourne, Victoria 3000, Australia, <sup>2</sup>Clinical & Health Sciences, University of Melbourne, Victoria 3010, Australia, <sup>3</sup>The Walter and Eliza Hall Institute of Medical Biology, University of Melbourne, Victoria 3010, Australia

#### Introduction

Despite the rapid expansion of both volume and complexity of proteomics data, the ability to easily leverage past or public experiments to enhance the interpretation of one's experimental results is still largely overlooked due to the computational complexity of the problem.

This step is instead often left to the researcher's own capabilities, knowledge and bias, potentially losing key

Mass Dynamics 2.0<sup>1</sup> (MD 2.0) unlocks this functionality by providing every life science researcher with straightforward access to state-of-the-art analysis techniques without the need to code and a unified place to

MD 2.0 enables simultaneous interrogation of multiple experiments

- Ability to access all or specific experiments to compare including all public experiments in app
- Interrogation of protein lists to deduce differential abundance behaviour across experiments • Interactive visualisation summaries of the significance across the selected experiments
- Collaboration and sharing functionality to improve communication and insight generation

### Case studies: How MD 2.0 facilitates multi-experiment functionality

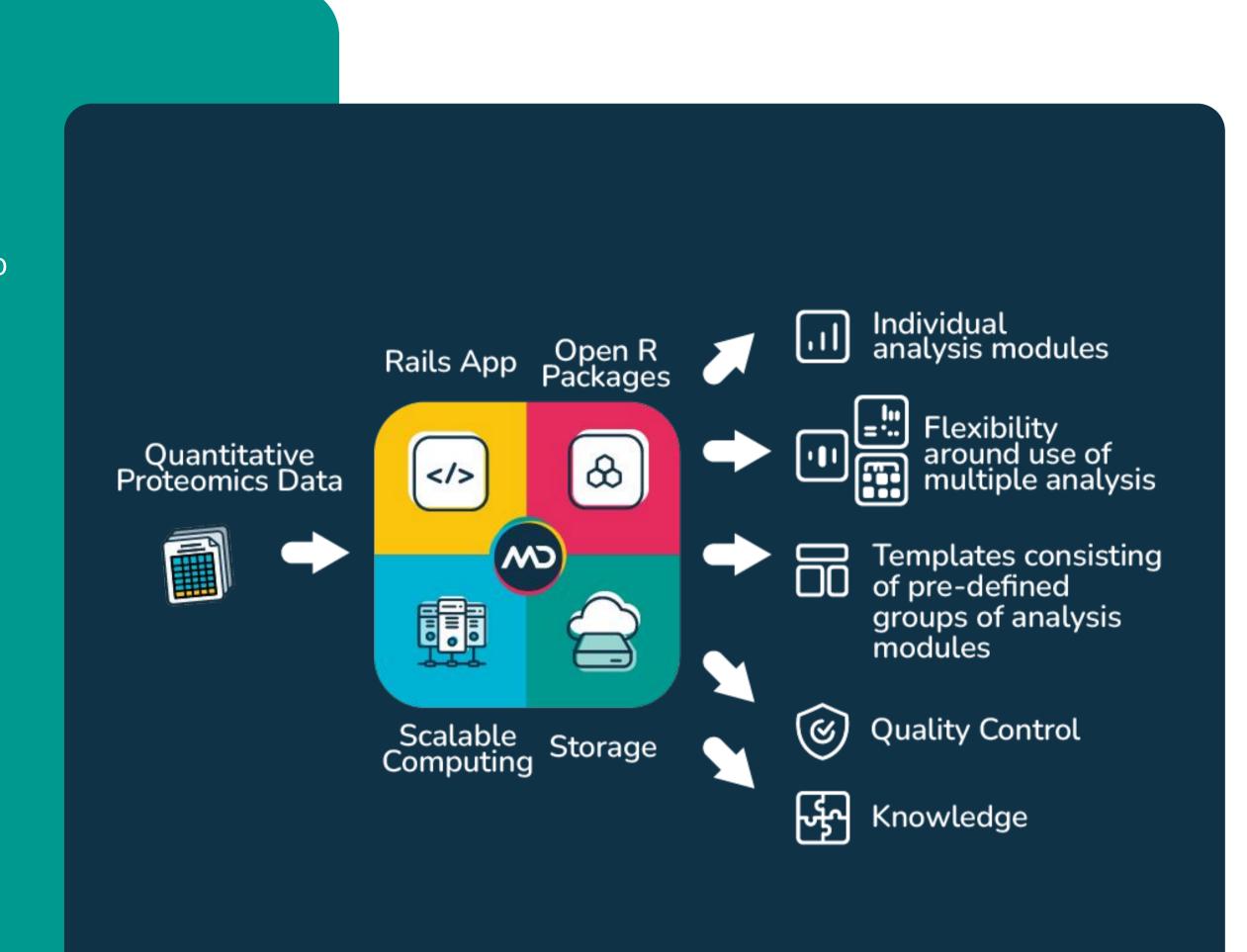
Here we present how MD2.0 facilitates proteomics analyses from data imports, processing and knowledge generation. The visualisations in this case study were created using the LFQ DDA datasets with PRIDE identifier PXD016433, PXD016447, and PXD019678<sup>2</sup>, containing respectively:

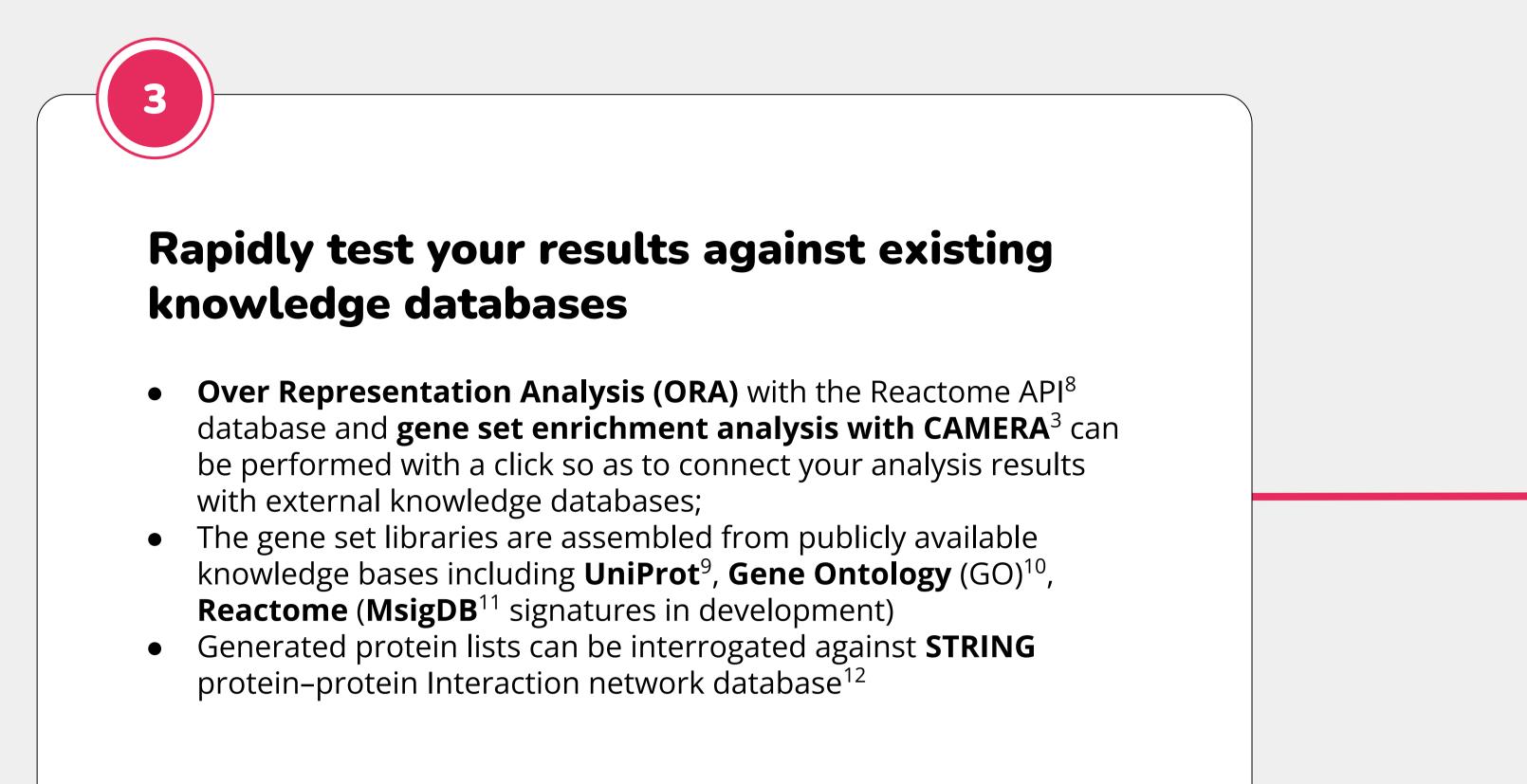
- 36 LFQ human urine samples (chronic kidney disease [CKD] stages 1, 3, and 5 vs healthy controls),
- LFQ analysis kidney tissue samples from a rat CKD model following filter-aided sample preparation (FASP), and
- Tandem mass tag (TMT)-labeled MS analysis of human primary glomerular endothelial cells (GECs) and proximal tubular epithelial cells (PTECs) before and after inducing 24-h hypoxia injury

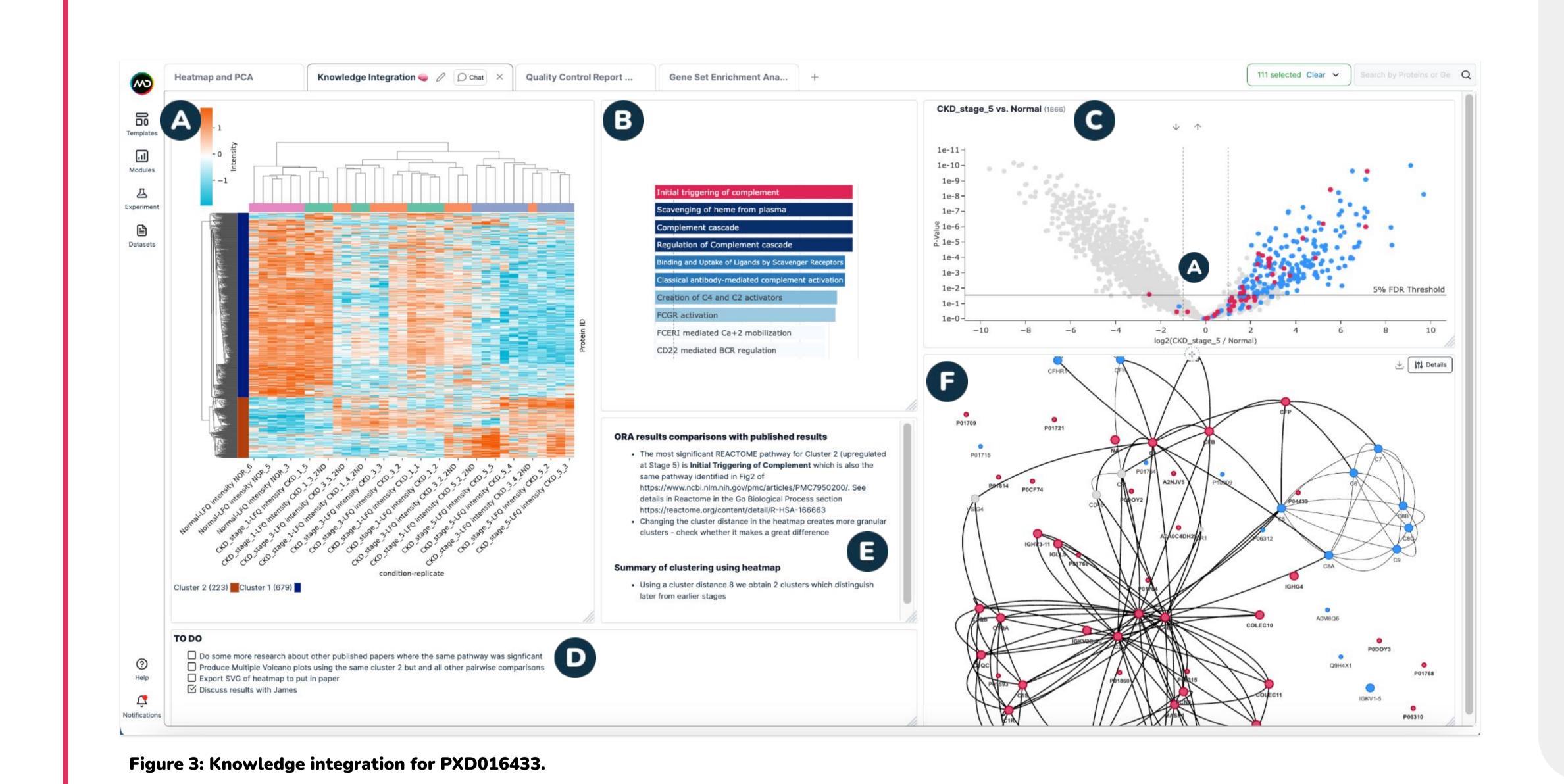
Accurately assess the quality of your experiment

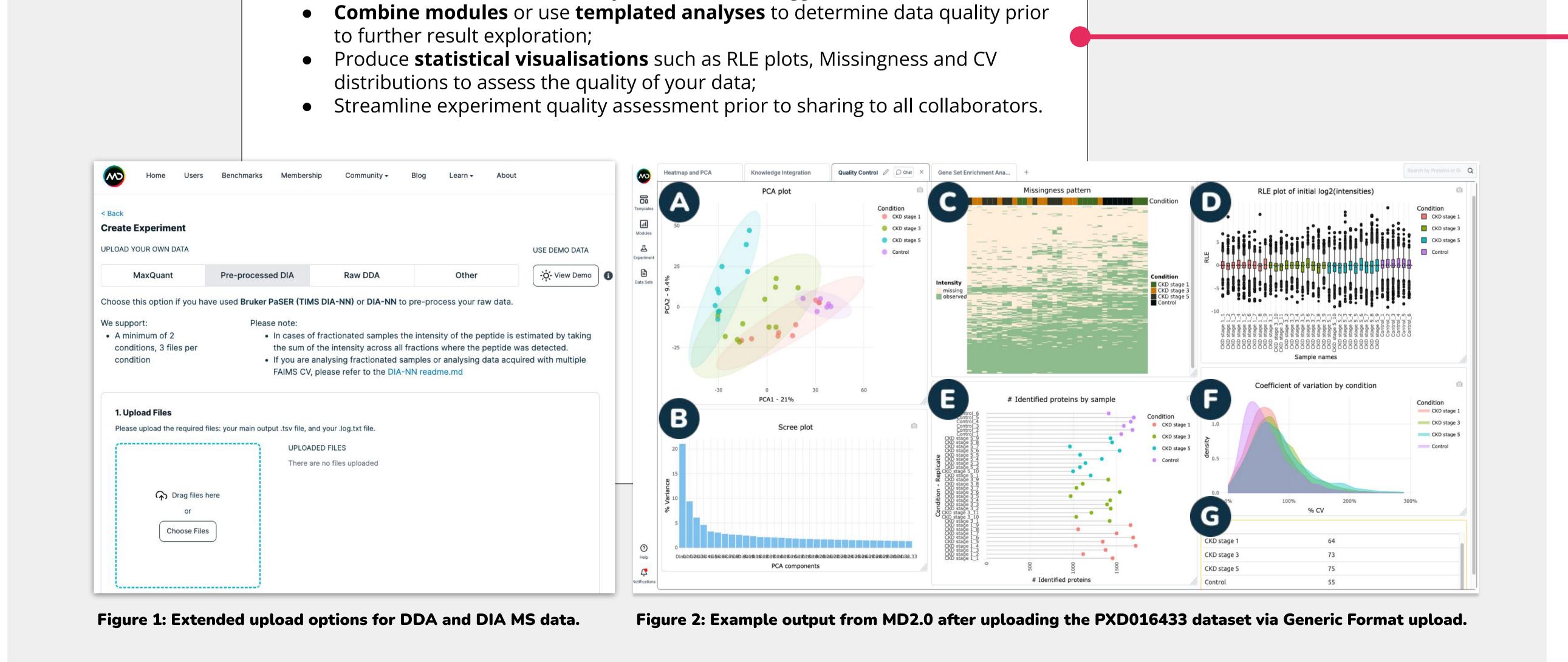
• **Upload** data from raw LC-MS output or from pre-processed analysis (e.g.

MaxQuant, **Bruker PaSER™ output**, DIA-NN, MSFragger);









# Interactively explore quantitative proteomics data:

- Directly run **state-of-the-art statistical methods** for differential expression and knowledge interpretation such as limma<sup>2</sup> and
- Work dynamically with alternate data visualisations like heatmaps,

Figure 2: Example output from MD2.0 after uploading the PXD016433 dataset via Generic Format upload. 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 table by condition. Figure 3: Knowledge Integration for PXD016433. (A) Heatmap to identify 2x main clusters. Cluster 2 (n=223) consisted of proteins that sequentially increased with increasing CKD severity and selected pathways such as complement activation, as previously described 1. (C) MD 2.0 allows users to link selected pathways. and their proteins in a pairwise comparison results with controls. (D, E) Mass Dynamics has dedicated check-list (D) and text (E) modules that allow you to setup checklists for you or your team and take notes/information around insights you have made from ORA... Figure 4. Workflow to leverage Multi Experiment Trend Analysis of consistently upregulated proteins in CKD compared with control; (B) Multi Experiment Trend Analysis allows interrogation of these across other experiment Trend Analysis allows interrogation of these across other experiment Trend Analysis allows interrogation of these across other experiments loaded into Mass Dynamics, using a fisher's combined probability test to combine p-values; (C) List selection and upload options; (D-E) Pairwise analysis and corresponding violin plots of PXD016433 (E), and the human protein IDs identified from PXD016447 (rat CKD model) with identified proteins from the trend analysis; (F-H) Corresponding human protein IDs and their abundance comparing CKD Stage 1 against control from PXD016433 (E), and the human protein IDs identified from PXD016437 (rat CKD model) with identified proteins from the trend analysis and corresponding human protein IDs and their abundance comparing CKD Stage 1 against control from PXD016433 (E), and the human protein IDs identified from PXD016437 (rat CKD model) with identified proteins from the trend analysis and corresponding human protein IDs and their abundance comparing CKD Stage 1 against control from PXD016433 (E), and the human protein IDs and their abundance comparing CKD Stage 1 against control from PXD016433 (E), and the human protein IDs and their abundance comparing CKD Stage 1 against control from PXD016433 (E), and the human protein IDs and their abundance comparing CKD Stage 1 against control from PXD016433 (E), and the human protein IDs and their abundance comparing CKD Stage 1 against control from PXD016433 (E), and the human protein IDs and the human protein ID Figure 5: Example taking notes, setting tasks, collaborating live with chat box and sharing experiment; (B) "Share" button to share experiment with collaborators (covered by the live chat in main panel).

### Explore differential protein abundance changes across multiple experiments uploaded to Mass Dynamics We can **compare protein abundance across conditions** in one experiment - e.g. generating protein lists that are significantly up in each stage of CKD relative to control in PXD016433. • Upset plots (Fig. 4A) can be adopted to select the consistent proteins that have higher abundance in CKD The Multi Experiment Trend Plot (Fig. 4B) module can be used to interrogate the selected proteins against other experiments previously uploaded (PXD016447 and PXD019678) Generated lists can be imported in various modules (Fig. 4C) for review/analysis • The Multi Experiment Trend Plot details allow you to select "All experiments" in your account, or allow you to specifically select experiments for interrogation Multiple settings define how to compare across experiments, including: Options on how to combine p-value method (e.g. Fisher's combined probability test) to provide an aggregate of statistical significance in one single visualization Filter significance options (p-value, adjusted p-value, none) Define significance threshold • Multi Experiment Trend plot leverages UniProt accession IDs, allowing the ability to search alternative IDs (e.g. secondary) + isoforms • For experiment comparisons across-species (Fig. 4D) it is possible to include UniRef similarity clustering to link protein IDs • Significant proteins identified in other experiments can be then imported to new experiments for analysis (Fig. G-H) Figure 4. Workflow to leverage Multi Experiment Trend Analysis of consistently differentially abundant proteins from PXD016433 across PXD016447 and PXD019678 PXD016447 - Rat Kidney CKD Model Quality Control Report Heatmap Pairwise Analysis Multi-exp Upset / Chat × GSEA SEA Multi-hypox-norm-hpTEC-cluster5 ✓ 20230522\_UP\_Stage1\_control Import a list from another experiment PXD016433 - Human Urine - LFQ Hypoxia-hPTEC vs. Normoxia-hPTEC (6918) Up-consistently\_across\_stg\_1-5 Combined p-value: 1.2936 Upload a list Import a list from another experiment ent: 20230512 PXD016447 CKD Demo data LFQ GenericFo COMBINE P-VALUE METHOD Fisher's combined probability test UNIREF CLUSTER SIMILARITY THRESH -10 -8 -6 -4 -2 0 2 4 6 8 10 log2(Hypoxia-hPTEC / Normoxia-hPTEC) Comments for Knowledge Integration 🧠 **Summary and Future Directions** Share, collaborate and publish, allow independent Mass Dynamics Multi-experiment trend analysis allows researchers to better leverage their other experiments to understand protein abundance trends in biological

#### 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; og2(CKD\_stage\_5 / Normal) It allows notes taking for improved 9 ₹ It allows export entire reports or specific earch about other published papers modules to \*.SVG as required; thway was signficant plcano plots using the same cluster 2 b Analysis of results in app can be made public to allow interactive assessment of tmap to put in paper results by reviewers and community.

## **Future Directions**

readilv available in a cloud environment.

#### Broaden upload options for various pre-processed outputs;

- Broaden statistical analyses options, e.g. time series and dose response analyses; Allow cross experiments comparisons;

• Possible because all data and results are stored in a standardised format and are

Uploading pre-existing datasets (e.g. from PRIDE) for trend analysis coupled with

external knowledge integrations improves interrogation of results for new insights

• Experiments are more readily searchable as a true internal knowledgebase

- Increase support to more knowledge bases (EnrichR, PRIDE, etc.);



