No One is Left Behind: Empowering Bioinformatics with Mass Dynamics Custom Workflow Capability

Anna Quaglieri[1], Aaron Triantafyllidis[1], Brendan Spinks[1], Mansi Aggarwal[1], **Mark R. Condina**[1,2], Paula Burton Ngov[1], Giuseppe Infusini[1] and Andrew I. Webb[1]
[1]Mass Dynamics, Melbourne, Victoria 3000, Australia, [2]Clinical & Health Sciences, University of South Australia, Adelaide, SA 5095, Australia

Introduction

The surge in complex proteomics experiments and the continuous development of analysis workflows at scale requires a move away from disconnected tools and manual processes to unlock new insights faster. **Mass Dynamics (MD)** exists to reduce barriers for users of all backgrounds, making established and cutting-edge methodologies accessible to all researchers in an intuitive, web-based collaborative environment[1].

After proteins or peptides are identified and quantified, quality control, statistical analyses, and validation steps are performed. **MD simplifies these processes by**:

- Providing an accessible, modular, web-based platform;
- Allowing seamless results comparison and multi-dataset management;
- Enabling researchers to organize and analyze their discovery and validation datasets in one place.

Here, we introduce MD's new **Custom Workflow** capability, designed for bioinformaticians to leverage their expertise directly on the platform, enabling any researcher to benefit directly from it to improve accessibility of internal bioinformatics workflow.

How can I integrate my Custom Workflow in the MD interface?

- Instructions to set up custom workflows in MD are publicly available on GitHub - https://github.com/MassDynamics/MDCustomR - with an end-to-end example in R that anyone can mimic. Similar concepts apply to Python workflows.
- MD can work with bioinformaticians in the team to support the deployment. Once the workflow is deployed, it will be easily accessible in the MD system for all members of the research team, with no need to code.

MD Custom workflow set up instructions

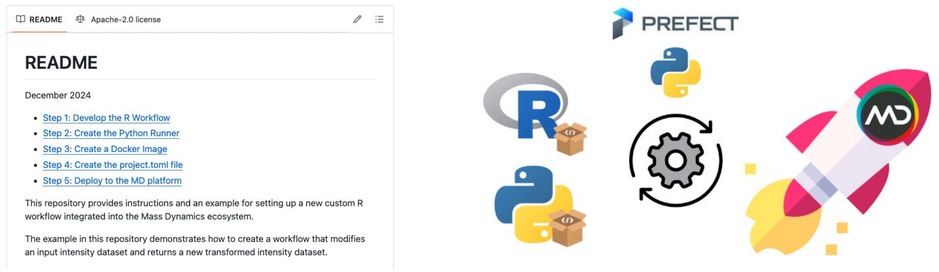

Figure 1. Setup instructions and overview of Custom Workflow capabilities in Mass Dynamics.

1 Develop your R Workflow

This is your R (Python) script containing your scientific workflow. The workflow can be **completely custom** as well as **proprietary** and can be **a simple script or a complex package**.

Our public example describes how to run an R package that can normalize a table of intensities using user-selected methods. Once your workflow is ready, you can set up an optional runner script to run the main function of the workflow (example in **Figure 2**).

```r
library(MDCustomR)

run_transform_intensities <- function(intensities, metadata, normMethod){

  print("Package Versions")
  print(packageVersion("MDCustomR"))

  output <- MDCustomR::transformIntensities(intensities = intensities,
                                            metadata = metadata,
                                            featureColname = "GroupId",
                                            replicateColname = "replicate",
                                            normMethod = normMethod)

  return(
    list(
      intensity=output$intensity, # required
      metadata=output$metadata, # required
      runtime_metadata=output$runtimeMetadata # optional
    )
  )
}
```

Figure 2. Example setup for the R runner to leverage the MD Custom Workflow capability. This script loads the `MDCustomR` package whose main function takes intensities and parameters in input and returns normalised protein intensities (https://github.com/MassDynamics/MDCustomR/blob/main/src/md_custom_r/process.R).

2 Create the Python Runner

This script (example in GitHub: https://github.com/MassDynamics/MDCustomR/blob/main/src/md_custom_r/process_r.py) leverages an open-source MD utility package to:

- Prepare the R input to be executed in Prefect - an open-source workflow orchestrator
- Define input arguments names and method descriptions that will be eventually displayed in the MD UI

3 Create a Docker image

Create a Dockerfile that contains all information and dependencies needed to install and run the workflow (example in GitHub: `Dockerfile`).

4 Create the project.toml file

An example is found in the GitHub page and only a few edits are required to provides the package description, including its versions, dependencies, and authors.

5 Deploy to Mass Dynamics

With the support of the team, deploy to Mass Dynamics and have your workflow integrated with the full system.

Internal test case: from script to user interface in half a day

Once the scientific R package was developed and ready to be adopted, it took one of our scientists and one engineer half a day to adapt the files in the example custom script repository and integrate a new workflow in the Mass Dynamics ecosystem.

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)

How is my Custom Workflow going to be used in Mass Dynamics?

After deployment, everyone in your team will be able to trigger and use your workflow directly from the Mass Dynamics interactive user interface.

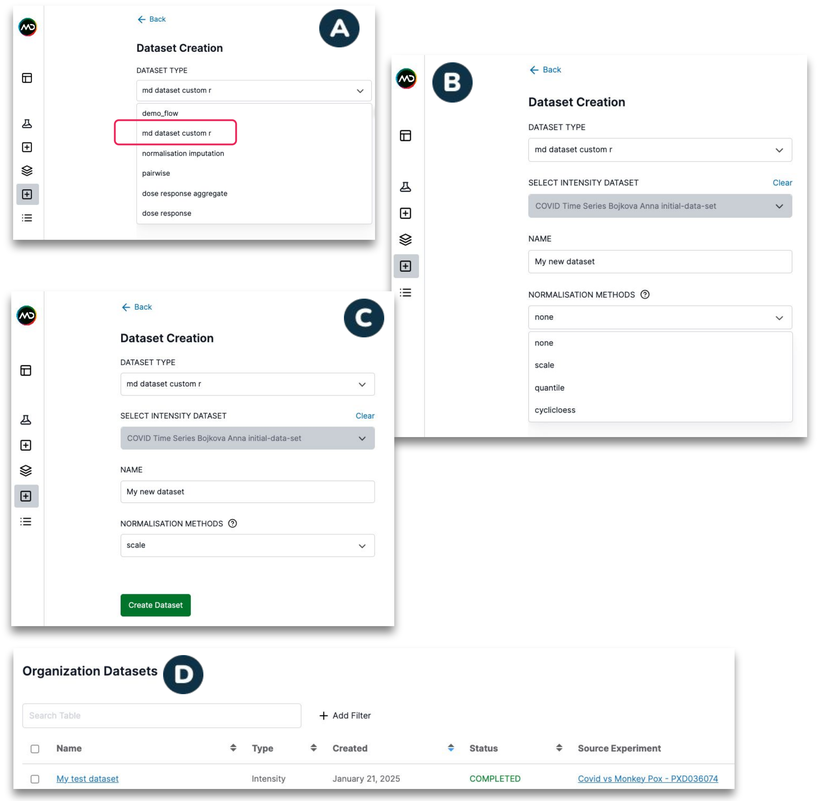

Figure 3. Display of how a new "md dataset custom r" workflow will be available in the MD user interface after following steps 1-5 in https://github.com/MassDynamics/MDCustomR/tree/main.

A. The new workflow will be available in the datasets dropdown, i.e. 'md dataset custom r'.
B. Input arguments and descriptions shown to the user as defined by the custom workflow developer in Step 2. See specifications for this example in https://github.com/MassDynamics/MDCustomR/blob/main/src/md_custom_r/process_r.py.
C. Once input data and arguments are specified, the new dataset can be created. In this example the normalisation method selected is "scale" to perform median normalization of the input intensities.
D. The processing status of the dataset creation is available in the Mass Dynamics user interface.

Your Custom Workflow results can now leverage the entire Mass Dynamics ecosystem

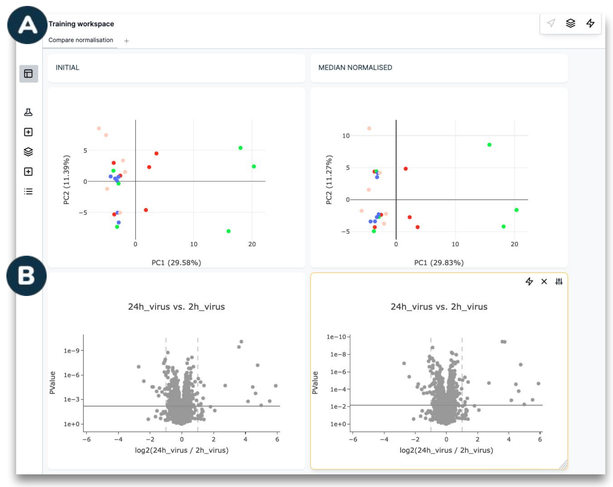

Figure 4. Leverage the MD ecosystem of visualisations to analyse your custom workflow results.

A. Once completed, the workflow output can be accessed and visualised using all available modules in the MD ecosystem. Here, the differences between the initial and median normalised intensities are compared next to each other using principal components analysis.
B. Side by side volcano plots compare the differential expression results with the initial and median normalised intensities.

Future Directions

The current publicly available example supports workflows to transform and output intensities tables. We are working to streamline the processing of any type of workflow with different inputs and output.

Ready to try Mass Dynamics?

Keen to try Mass Dynamics using your own data? Simply click on the QR code to book a custom demo.

Conflicts of interest

The authors Aaron Triantafyllidis, Paula Burton Ngov, Giuseppe Infusini and Andrew I. Webb declare that 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, Brendan Spinks, Mansi Aggarwal and Anna Quaglieri are employees of Mass Dynamics.