

# NO ONE IS LEFT BEHIND: EMPOWERING BIOINFORMATICIANS WITH MD 3.0'S CUSTOM WORKFLOW CAPABILITY

### 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<sup>1</sup>.

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 replicate. 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.





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

### 

#### **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).



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).

#### References

1. Quaglieri A, et al. Mass Dynamics 2.0: An improved modular web-based platform for accelerated proteomics of SARS-CoV-2-infected host cells reveals therapy targets. Nature (2020)

#### **Create the Python Runner**

This script leverages the open-source MD utility md dataset package

- A. Define input arguments names and method descriptions that will be eventually displayed in the Mass Dynamics UI
- 3. Prepare the R input to be executed in **Prefect** - an open-source workflow orchestrator



Figure 3. Example setup for the Python runner. Working example in GitHub

https://github.com/MassDynamics/MDCustomR/blob/main/src/md\_cu <u>stom\_r/process\_r.py</u>.

#### Create a Docker image

Create a Dockerfile that contains all information and dependencies needed to install and run the workflow (example in GitHub: https://github.com/MassDynamics/md\_dataset/blo <u>b/main/Dockerfile</u>).

# 4

#### Create the project.toml file

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



3

#### **Deploy to Mass Dynamics**

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

**MD** Custom Workflow set up instructions



\_\_\_\_\_



#### 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.



# READY TO TRY MASS DYNAMCS?

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

# 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.

	← Back		← Back			← Back	
	Dataset Creation		Dataset Creation			Dataset Creation	
	DATASET TYPE		DATASET TYPE				
	md dataset custom r		md dataset custom r	~		DATASET TYPE	
	demo_flow		SELECT INTENSITY DATASET	Clear		md dataset custom r	~
凸	md dataset custom r	2	COVID Time Series Bojkova Anna initial-data-set	×	凸	SELECT INTENSITY DATASET	Clear
+	normalisation imputation	Ð			Đ	COVID Time Series Bojkova Anna initial-data-set	~
	pairwise	8	NAME My new dataset				
÷	dose response aggregate	Ŧ				NAME	
	dose response	:=	NORMALISATION METHODS ⑦		Ŧ	My new dataset	
:=			none	~	:=	NORMALISATION METHODS ③	
			none			scale	~
-	splay of how a new "md dataset custom r"		scale				
	vill be available in the MD user interface after		quantile				
<b>following steps 1-5 in</b> https://github.com/MassDynamics/MDCustomR/tree/main.			cyclicloess			Create Dataset	
	workflow will be available in the datasets dropdown, i.e. set custom r'.						
	uments and descriptions shown to the user as defined by						
	m workflow developer in Step 2. See specifications for this	Organization	Datasets			New Dataset	
example i	:hub.com/MassDynamics/MDCustomR/blob/main/src/md_c	1 State 1 Stat					
example i <u>https://git</u> <u>ustom_r/p</u>	<u>:hub.com/MassDynamics/MDCustomR/blob/main/src/md_c</u> process_r.py.	Search Table	+ Add Filter				
example i <u>https://git</u> <u>ustom_r/p</u> <b>C.</b> Once inpu	hub.com/MassDynamics/MDCustomR/blob/main/src/md_c	Search Table	+ Add Filter	Status	In workspace	Source Experiment	

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

**Data:** The data used is a COVID dataset with protein measurements taken from Virus and Control cells at different time points (Bojkova et al,  $2020)^2$ .



#### using all available modules in the MD ( Here, the differences between the initia median normalised intensities are com to each other using: **A-B**. Principal com analysis and C-D. Intensities distribution boxplots of Relative-Log Expression (F The boxplots confirm that after median normalisation the distributions median Figure 5. Compariso (DA) analysis before

A. Log-Log plot of th median normalise from the diagona whose logFC cha such protein, PRL protein is highligh plot. B. Volcano plot comp initial intensities h

significance. C. Volcano plot comp after median norm

## **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.

# **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.

Anna Quaglieri<sup>1</sup>, Aaron Triantafyllidis<sup>1</sup>, Brendan Spinks<sup>1</sup>, Mansi Aggarwal<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</sup> <sup>1</sup>Mass Dynamics, Melbourne, Victoria 3000, Australia, <sup>2</sup>Clinical & Health Sciences, University of South Australia, Adelaide, SA 5095, Australia

alised ystem. d		A			MEDIAN NORMALISED
next its ig lues.	⊥ ÷ ⊗	6- (spin <sup>-</sup> 4- - - - - - - - - - - - - - - - - - -	B 1e-10- 1e-8-	24h_virus vs. 2h_virus	24h_virus vs. 2h_virus
<b>bifferential Abundance</b> <b>after normalisation.</b> comparisons (x-axis is the aset, y-axis initia). Shifts ity line show proteins after normalisation. One shighlighted. The same h each separate volcano two conditions using the hting the protein changing the same two conditions tion.	i≡ ⊘	PRUNE2: Highlighted is a protein that changes fro not differentially abundant to differentially abund		PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2 PRUNE2	$1e^{-8}$ $1e^{-6}$ $1e^{-4}$ $1e^{-2}$ $1e^{-4}$ $1e^{$