An ACF Multicentre Testing Initiative: Exposing the Limitations of Shotgun Proteomics

Erwin Tanuwidjaya[1], Giuseppe Infusini[2], Mark Condina[2], Ralf B. Schittenhelm[1]

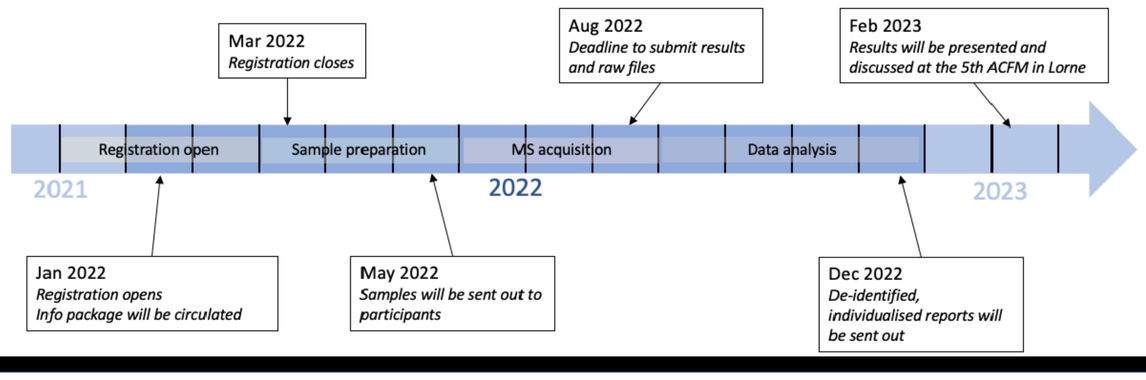
1 – Monash Proteomics & Metabolomics Facility, Monash Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia

2 – Mass Dynamics, Melbourne, VIC, Australia

INTRODUCTION

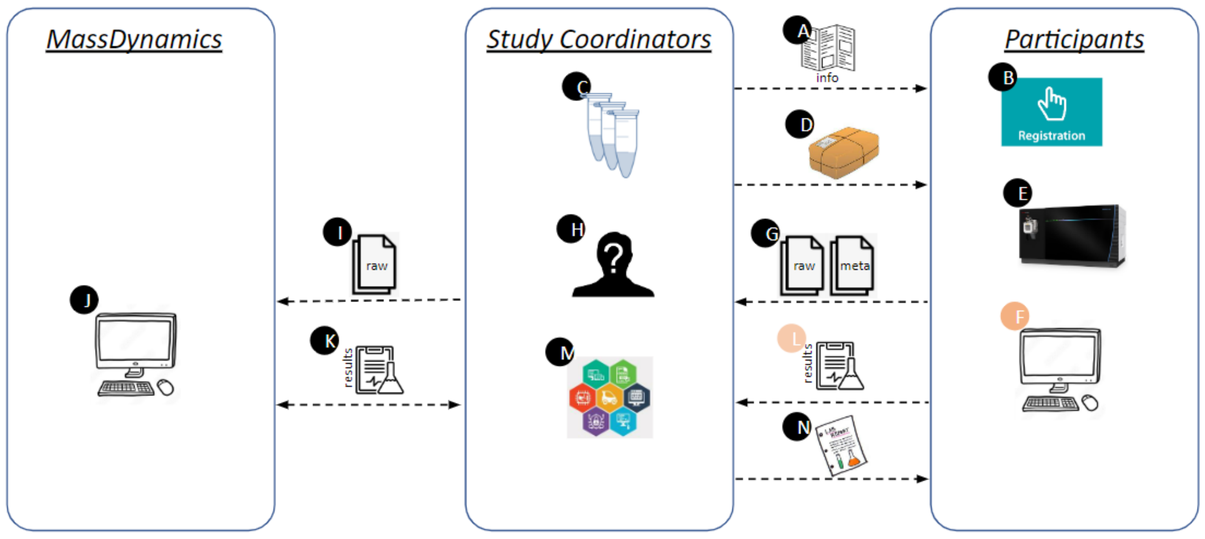
- The Multicentre Testing Initiative (MTI) is an Australasian Core Facilities (ACF) initiative which aims to **examine reproducibility across participating facilities as well as to identify optimal strategies and workflows for specific research questions and/or sample types**.
- To achieve this, a series of well-defined, independent studies will be conducted across participating facilities that target various mass spectrometric areas and techniques.
- As such, the MTI will provide participants with an opportunity to anonymously evaluate their in-house workflows in comparison to others.

Timeline

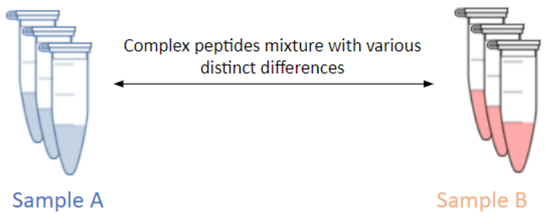


AIM:

MTI – Study 1 **specifically focuses on strategies and workflows to determine relative proteomic difference across two samples using data-dependent acquisition (DDA) mass spectrometry (MS).**



METHODS



1. All samples were centrally prepared by Monash Proteomics & Metabolomics Facility (MPMF).
2. Participants were not informed of the sample composition.
3. A total of **5** proteins and **11** iRT peptides of different concentrations were spiked into K562 whole cell extract.
4. Lyophilized samples (30ug) were shipped to participants alongside sample reconstitution guidelines and FASTA database.

Sample acquisition guidelines:

Any hardware setup capable of LC-MS/MS analyses and any workflow/method can be used.

Following restrictions apply:

- ❑ DDA mode (shotgun proteomics; IDA)
- ❑ Linear separation gradient ≤120 min
- ❑ No pre-fractionation
- ❑ Per dataset: Technical triplicates (n = 3) → 6 injections
- ❑ Capped injected amount at 2 ug per replicate (nanoflow)

Raw files were to be submitted to MPMF along with metadata and bioinformatics forms.

Participants are strongly encouraged to analyze the samples on **as many mass spectrometers with as many methods as possible**.

METADATA

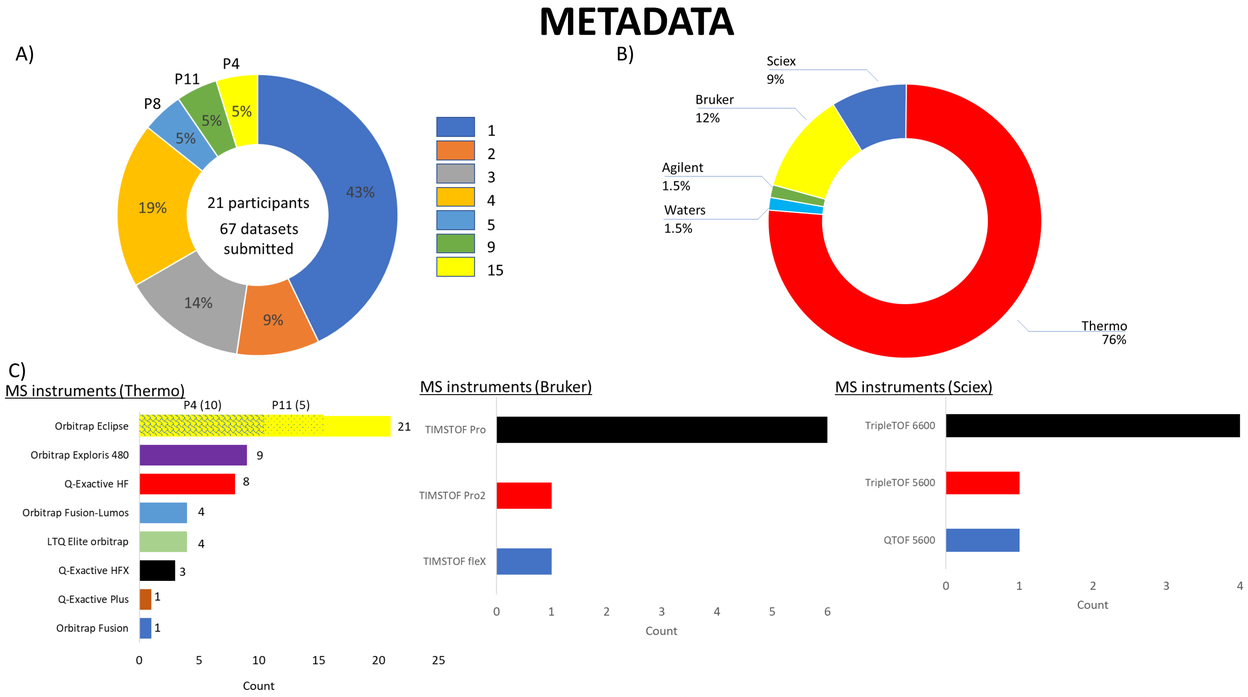


Figure 1. Information of the participants and mass spectrometers in MTI – Study 1. A) Breakdown of the participants and datasets submitted. **B)** Breakdown of mass spectrometers by brand, and **C)** breakdown of mass spectrometers by instruments.

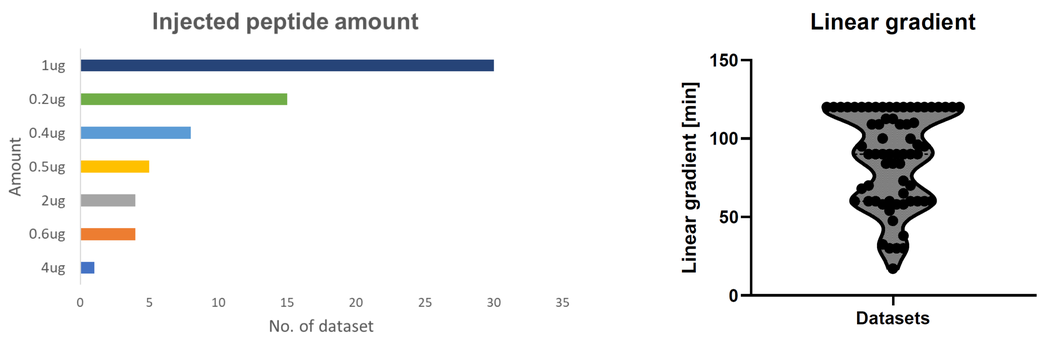


Figure 2. Distribution on pre-acquisition parameters. A) Injected peptide amount on column, and **B)** Linear gradient on which the acquisition is ran.

RESULTS

Proteins

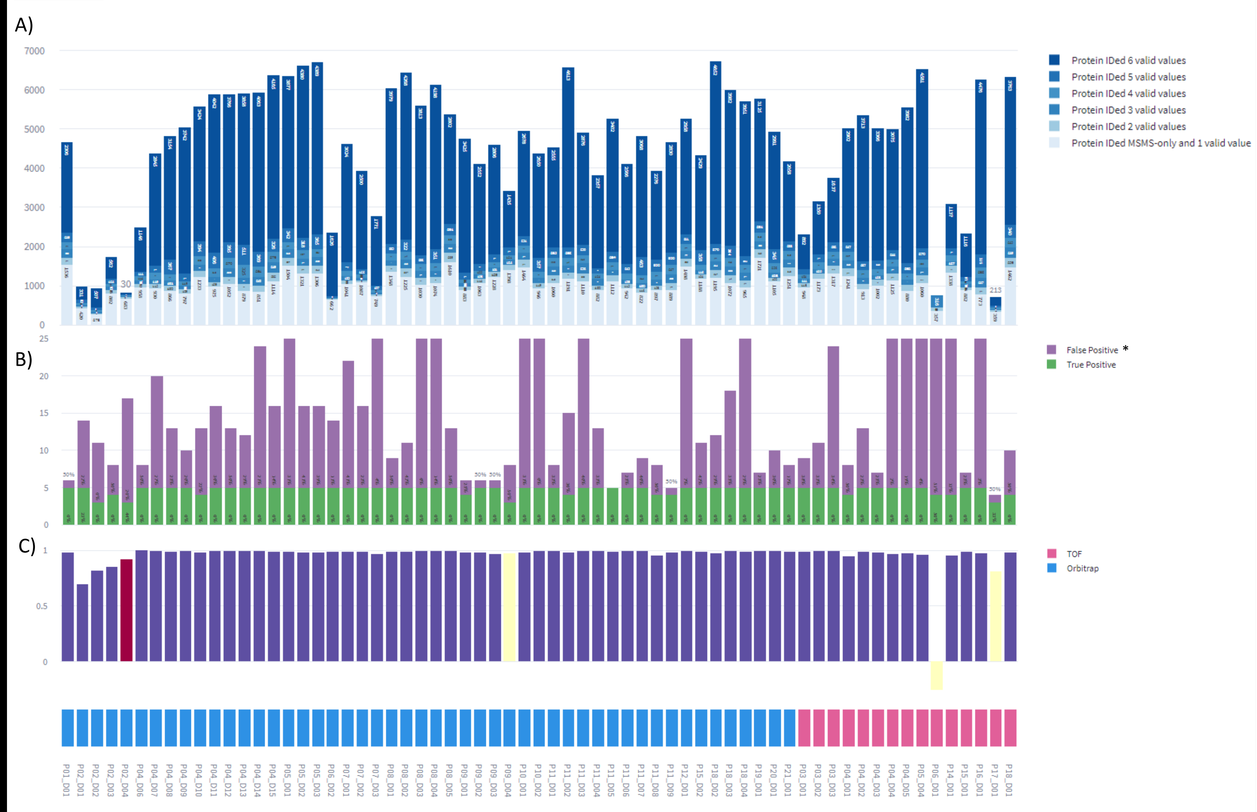


Figure 3. Bar charts illustrating the protein spike-in identification and quantification, sorted by MS types. A) Protein identification of all datasets, B) Number of significant differentially expressed proteins identified between Sample A and Sample B, and C) Estimated ratio accuracy between theoretical and identified protein spike-ins.*false positive values extend beyond 25 shown in B). The maximum value is at range is between **25 to 1502** (P16_D01).

Peptides

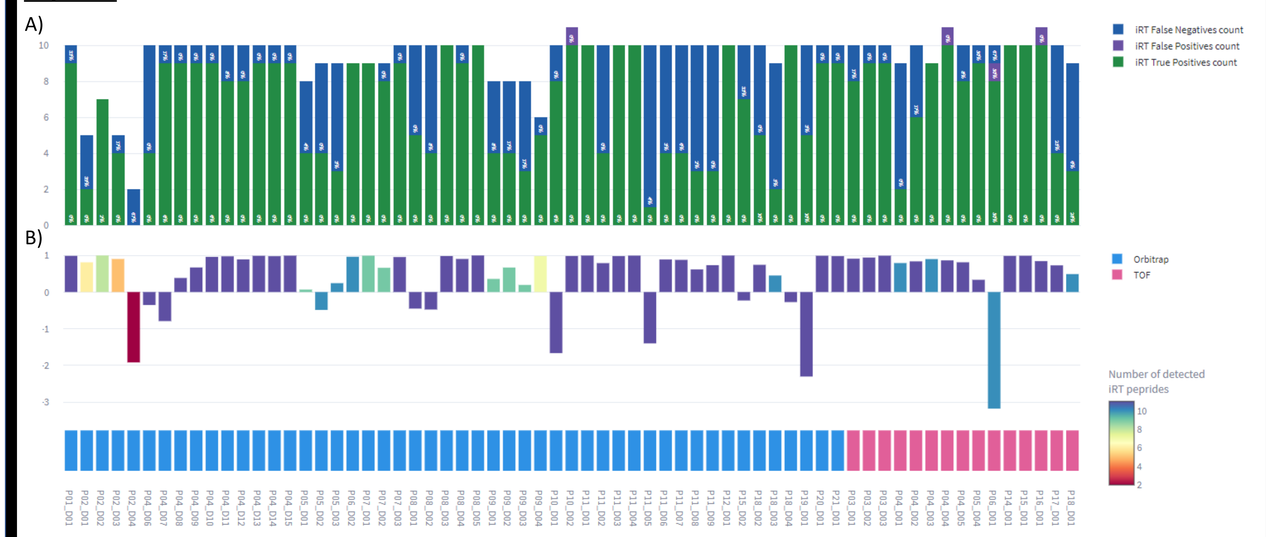


Figure 4. Bar charts illustrating the peptide spike-in identification and quantification, sorted by MS types. A) Number of significant iRT peptides identified between Sample A and Sample B, and B) Estimated ratio accuracy between theoretical and identified iRT peptide spike-ins.

CONCLUSION

1. There are great variations observed between different pipelines, shown by the difference in the number of IDs found at the protein identification level.
2. Majority of the participating facilities shows high accuracy in identifying and quantifying the spike-in proteins, indicating a strong confidence in proteomics facilities' technical expertise.
3. Despite facilities ability to identify true positives, there are significant variance in the number of false positives highlighting the fact that DDA severely overestimates the number of significantly regulated proteins, presumably due to imputation strategies.
4. A more complex proteomics study with potential clinical significance may be of benefits as an extension to MTI – Study 1.

ACKNOWLEDGEMENT

We thank the participating proteomics facilities for the time and commitment to this study. We would also thank the Mass Dynamics Team in developing the interactive analytical web application which allows participants and the broader proteomic community to optimise their own proteomic workflow.

WORK IN PROGRESS

1. Further analysis of both metadata as well as the identifications of proteins and peptides
2. Finalization of Interactive analytical web application (by Mass Dynamics Team)