

Automated benchmarking and quality control for protein characterisation workflows allows rapid development and fast confidence

Authors

[Anna Quagliari](#)¹; Joseph Isaac Bloom¹; Susmita Saha¹; Brendan Spinks¹; Aaron Triantafyllidis¹

[View Session Details](#)

Institutes

¹Mass Dynamics, Melbourne, Australia

Introduction

Protein characterization, including peptide mapping and multi-attribute monitoring using LC-MS/MS methods are widely adopted techniques in both academic and industrial drug development settings. Holistic peptide mapping workflows are computationally complex, involving many algorithmic steps, usually requiring manual input and decision making from the end user and therefore extra effort to guarantee reproducibility of the results. While automated workflows are increasing in popularity, there is a need to make sure that ease of use doesn't come at the expense of quality and transparency. Here, we present an automated and standardised approach to benchmark and quality control protein characterisation workflows.

Methods

The Mass Dynamics Peptide Mapping service is a fully automated, cloud-based LFQ-DDA workflow that presents identified and quantified peptide spectral matches via an easy to use web interface. To guarantee the ongoing accuracy and appropriateness of all algorithmic steps including Feature Detection, Database Search, Target-Decoy Scoring, Mass Recalibration and Targeted Feature Extraction, we developed automatic and interactive quality control reports generated alongside experiment results. Building on this work, we developed an automated benchmarking service using publicly available data with ground truth that can validate and evaluate each new release of the workflow.

Preliminary Data

The quality control report provides plots and summary statistics for each step in the workflow. This includes for example the number of features detected in each sample; the cross correlation distribution between target and decoy after peptide scoring; the mass over charge error distribution after performing MS1 recalibration and the number of high scoring identifications detected in one or multiple samples. Detailed documentation of the interpretation and expected behaviour is provided for each reported statistics and this will allow quick comparison between experiments and an increased understanding of how to detect outlier samples and characterise the quality of an experiment. For example, while the number of features detected in each sample isn't directly associated with quality, it can be used to compare different runs of the same experiment to highlight outlier samples. Moreover, the cross correlation distribution between target and decoy peptides should have a long right tail for target peptides and the mass over charge error should be centered around zero after recalibration. The automated benchmarking service uses publicly available data with known concentrations of spiked proteins. Benchmarking statistics are provided such as the number of peptide spectrum matches identified in each replicate of the same sample or comparison of peptide quantification results with ground-truth ratios. Benchmarks are run with every new release of the workflow and are available to the user for inspection. The automated benchmarking setup allows the user to maintain trust and reliability in the service as well as it allows a robust and efficient development and testing of the workflow by providing real-time results using ground truth data.

Novel Aspect

Standardised and automated quality control and benchmarking tool for protein characterization.

