

## **Reproducible multi-omics workflows to get the most out of your data**

Reproducible workflows are crucial in all scientific fields, including the analysis of bioinformatics data. I will highlight some community efforts that aim to facilitate these, to which I have contributed. In proteomics, SDRF has facilitated analyses by providing a computer-readable format. ProteoBench on the other hand enables the direct comparison of different workflows on a standard datasets, assessing the comparability of various methods. Quality control of the data is a vital aspect prior to conducting any analyses. To address this issue, the mzQC community project aims to improve the exchange and communication of QC metrics.

Having covered these foundational concepts, I will delve into some of the methods that I have recently been involved in developing. To make the most of proteomics data, it is sometimes necessary to include non-canonical sequences in the analysis. ProtGraph is a method that applies all known sequence information from sequence databases and represents varied annotations in a graph, thus avoiding combinatorial explosion. This allows peptides that fit spectrum precursors to be quickly exported into FASTA for PSM matching by any current identification workflow.

When genomics and/or transcriptomics data are available, we have developed a workflow that transforms reads into the respective peptide sequences. This enriches the proteome of the sample species with sequences found in the reads, including mutations and alternative splicing. However, this can result in large sequence databases, which pose challenges for several search engines, particularly in meta-proteomics and immuno-peptidomics analyses. To address this issue, we are developing MaCcoyS (Mass Centric Decoy Search), which uses peptides stored in McPepDB for rapid searches. It also aims to improve FDR estimation using a spectrum-based approach.

I will highlight these ongoing projects and demonstrate how multi-omics approaches can provide a deeper exploration of the respective data.