University of Zurich[UZH] ETH zürich

f·g·c·z functional genomics center zurich

prophosqua

A Workflow to Analyse Phospho Proteomics Data

Jonas Grossmann, Witold Wolski, Antje Dittmann, Leonardo Schwarz, Ralph Schlapbach and Christian Panse

More info: protinf@fgcz.uzh.ch

SCAN ME

Introduction

Phosphoproteomics or other PTM-studies are vital for understanding many different processes in molecular biology. Labeling peptides with isobaric mass tags (e.g. TMT-labels) and subsequent multiplexing, is particularly beneficial . It enhances sensitivity as it is highly compatible with enrichment and fractionation methods and typically does not suffer from missing data. Pre-enrichment and fractionation are often necessary to analyze and achieve sufficient sensitivity for the moiety of interest (e.g., phospho-peptides). However, to study a differentially expressed PTM-site effectively, analyzing the total proteome is crucial. This allows us to differentiate the observed changes of the phospho-site from protein abundance changes. As a core facility, we prioritize flexible solutions that deliver good results in a reasonable time and, ideally, can be automated. Our proposed R-package meets these criteria, offering flexibility in its use and the potential for automation, thereby streamlining and accelerating PTM-studies.

Method

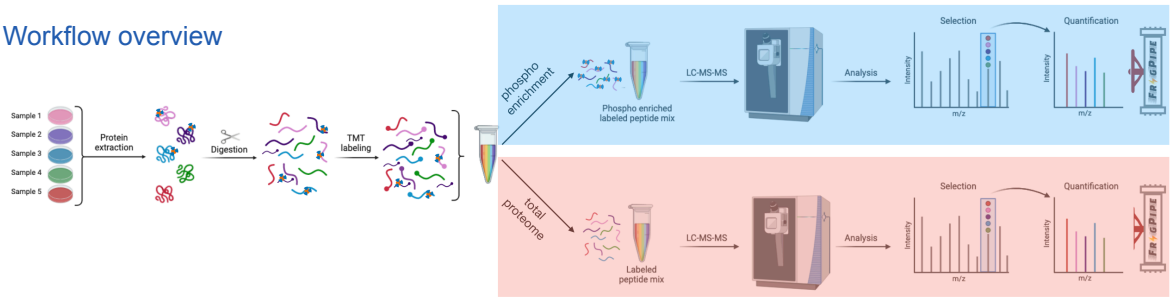
Here, we present an R-package prophosqua that extends the functionality of prolfqua [1] and prolfquapp [3], to analyze PTM-sites and allow it to be combined with the analysis of the total proteome. To ensure the reliability of our tool, we have benchmarked phrophosqua using a dataset with simulated data, comparing its performance thoroughly to MSstatsPTM[4].

Availability

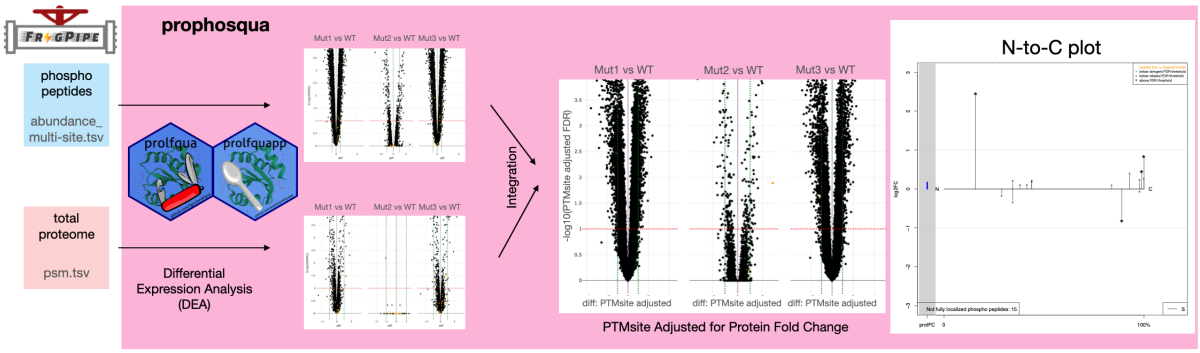
https://github.com/prolfqua/prophosqua

Poster

Workflow overview



Wet-lab workflow: For multiple samples, proteins are individually extracted, digested, and chemically labeled. Afterward, all samples are pooled, followed by high pH reverse phase offline fractionation and a phospho-enrichment step. Enriched phospho-peptides are measured and analyzed using FragPipe (v.21.1). The flow-through of the enrichment step (not phosphorylated peptides) is also measured and analyzed similarly with appropriate settings (referring here as total proteome).



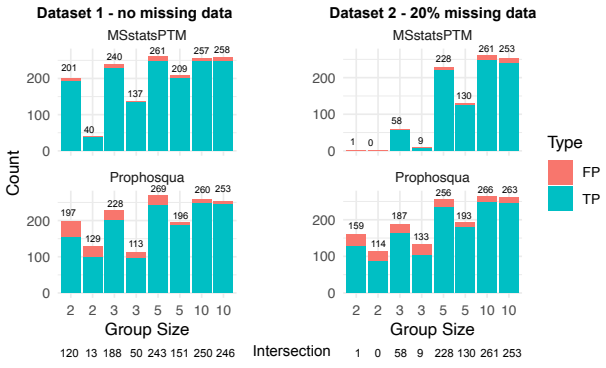
Analysis workflow: In prolfquapp, we provide R scripts to analyze the "psm.tsv" file for the total proteome, giving us the flexibility to define normalization and roll-up procedures. For the phospho-enriched part, we use the "abundance_multi_site.tsv". First, an individual differential expression analysis (DEA) is performed prolfqua and prolfquapp. Finally, we integrate the individual DEA results, which allows us to adjust each identified PTM site using the change estimated for the protein. Furthermore, we visualize these PTM-sites, e.g., directly on the protein's backbone, indicating the different log2FC for the different phospho-sites found for this protein along with the protein log2FC.

Performance

MSstatsPTM simulation datasets:

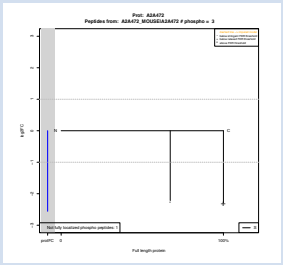
- https://github.com/devonjkohler/MSstatsPTM_simulations/
- 2 conditions are compared
- Different number of replicates per condition
- ***Phospho-enriched:*** For each condition 2000 modified peptides from 1000 proteins (two per protein) are simulated, 1000 (500 proteins) without change, 1000 (500 proteins) with a simulated difference of ~ 1.7 fold between conditions
- ***Total proteome:*** 1000 proteins were simulated each 10 peptides for 250 proteins the same difference was introduced as for the PTM peptides while 750 are simulated without change between conditions
- Dataset 1: all complete, no missing data
- Dataset 2: 20% of all the features missing at random
- Evaluated are significant PTM-sites with FDR (BH) < 0.05
- Ground-truth (TP = 250): all PTM-sites where only the PTM-peptide is changing and NOT the protein. All other PTM-sites (FP = 750) are considered to be FP

Prophosqua vs MSstatsPTM



Dataset 1: The number of significant calls is very similar for prophosqua and MSstatsPTM. However, for small group sizes (n=2,3), MSstatsPTM is more specific and reports fewer FP.

Dataset 2: For small group sizes (n=2,3), MSstatsPTM fails to make significant calls in the presence of missing observations. However, prophosqua can make many significant TP calls, although the FDR is slightly underestimated.



The N-to-C plot shows log2FC of PTM-sites (before adjusting for the protein change) along with the observed change for the protein. In this example here, the protein change mainly explains the observed change on the PTM-sites. These phospho-sites will not reported as significant when adjusting with the protein fold change.

References:

[1] prolfqua: Wolski W, et. al. prolfqua: A Comprehensive R-Package for Proteomics Differential Expression Analysis. JPR 2023 April; doi: 10.1021/acs.jproteome.2c00441

[2] prolfquapp: https://github.com/prolfqua/prolfquapp

[3] MSstatsPTM: Kohler D, et. al. MSstatsPTM: Statistical Relative Quantification of Posttranslational Modifications in Bottom-Up Mass Spectrometry-Based Proteomics. Mol Cell Proteomics. 2023 Jan;22(1):100477. doi: 10.1016/j.mcpro.2022.100477

Acknowledgement:

The authors would like to thank all the FGCZ staff and the FGCZ community for constant positive interactions and discussions about all various topics.