BIOGNOSYS
NEXT GENERATION PROTEOMICS

Biognosys AG
Wagistrasse 21
8952 Schlieren (Zurich)
Switzerland

t: +41 (0) 44 738 20 40
f: +41 (0) 44 738 20 49
info@biognosys.com
www.biognosyscom

A Systematic Characterization of LC-MS Features Sheds Light on the Full Potential of DIA Identification

Abstract ID number: 314026

Presenter: ***Grzegorz Skoraczynski***

Introduction

Comprehensive proteome identification coverage remains crucial for a better understanding of the mechanisms of diseases. Currently, in proteomics, the experimental method of choice is liquid chromatography-mass spectrometry (LC-MS) with data-independent acquisition (DIA). Over recent years, continuously improved identification algorithms have largely increased the number of identifications in DIA. Still, however, significant fractions of samples remain not identified. The goal of this project is to estimate how much more we can identify from the DIA method and how these improvements may be realized. To this end, we did a bioinformatics analysis of 4 systematic datasets to characterize the DIA feature space.

Methods

In this work, we wanted to find the characteristics of the non-identified signal. We analyzed 4 systematic datasets of HeLa cell lines, for which either gradient length or sample loading was varied. All datasets were acquired on 2 different instruments and acquisition types: Thermo Fisher Orbitrap and Bruker timsTOF.
Every dataset was processed using a library-free identification workflow directDIA+ from a modified version of Spectronaut 17 software capable of exporting all detected features. Then, we mapped these features with corresponding identifications to explore the space of non-identified and identified signals. Particularly, we also investigated the proportion of identified signals and how they are distributed.

Preliminary Data

Analysis revealed complex and non-homogenous patterns of peptide identification, relative to dataset type, experiment type as well as analyzed metrics. We observed that for both instruments, increasing the gradient length led to more detected features (on average 3.5 times more from shortest to longest gradient) as well as improved the ratio of identified features (on average by 15 % from shortest to longest gradient). However, for loading amount, we observed that while higher loading led to more detected features in both datasets (on average 3.4 times more features), the ratio of identified features increased by 12 percentage points with lower loading amounts in timsTOF.

Biognosys AG
Wagistrasse 21
8952 Schlieren (Zurich)
Switzerland

t: +41 (0) 44 738 20 40
f: +41 (0) 44 738 20 49
info@biognosys.com
www.biognosyscom

We also observed that in all cases the ratio of identified signals increased with increasing quality of the features. But even in the top 10 percentiles of features, over 30 % of features usually were not identified. We hypothesize that a fraction of these features is likely due to not having the peptide in the search space. To see how many of these peptides may be modified, we used FragPipe v. 19.1 with an open search and we identified about 7.5 % more features (8.8 % more in the top 10 percentiles).

Concluding, we characterized the areas where identification algorithms fail and suggested how to improve them and thus enabling deeper proteome identification coverage.

Novel aspect

The in-depth analysis of feature and identification characteristics allows for the removal of bottlenecks of protein identification.