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

Improved DIA-PASEF Based Quantitative Proteomics Using Spectronaut

Sira Echevarria-Zomeno, Tejas Gandhi, Oliver M. Bernhardt, Normand Overney, Dariush Mollet, Anna Susmelj, Lukas Reiter

Introduction

Ion mobility (IM)-based data independent acquisition (DIA) can bring improvements by an additional dimension of ion separation and a boost in sensitivity. We have previously presented our dia-PASEF analysis pipeline with high-precision IM (Gandhi, ASMS 2020) which automatically defines the optimal extraction window per peptide in IM dimension. Since then, we have greatly improved Spectronaut via a new machine learning framework (Overney, ASMS 2021) and deep learning-based scores. In this work we evaluated the impact of these efforts on the processing of dia-PASEF data with a 4- species controlled quantitative experiment. We found a significant improvement in not only overall identification rate, but also in quantification.

Methods

We prepared a 4-species mixed proteome sample with two conditions (H. sapiens 1:1, S. cerevisiae 1:2, E. coli, 1:10, C. elegans 1.3:1). A 2h gradient was used. For library generation, the pooled sample was fractionated and acquired with a timsTOF Pro in PASEF mode. Using Pulsar, our database search engine, the library resulted in 194,515 precursors and 19,537 proteins across all species. For the quantitative experiment, each condition was acquired in triplicates in dia-PASEF mode. The targeted analysis of dia- PASEF runs using both Spectronaut v15 and v16 with 1% FDR at peptide and protein level.

Results

Spectronaut v16 identified ~14,000 protein groups cumulatively across the four species, an improvement of 33% over Spectronaut v15. When considering only identifications with CV < 20%, v16 still identified 17% more protein groups on average for both conditions. Both versions performed similar in accuracy despite having more identifications with v16. The median absolute deviation from the expected log ratio between two conditions across all 4 species was 0.07 in v15 and 0.11 in v16. Furthermore, we performed regulation analysis using an unpaired t-test at protein level (Huang 2019). Upon sorting all the regulated candidate pairs by p-value, v16 recovered 9.3% more true candidate pairs (3046 candidates) with a 5% false positive rate. Finally, we used additional dia-PASEF datasets covering different sample types and gradient lengths to benchmark the overall identification performance between the two versions. For v16, we saw an average improvement of 44% at precursor and 25% at protein group level with ultra-short gradient (≤ 20 mins) and of 16% at precursor and 5% at protein group level for longer

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

gradient (≥ 60 mins) dia-PASEF datasets over v15. We speculate that the reason for a relatively stronger improvement in the four species experiment and the samples acquired with ultra-short gradient is due to the higher complexity of the data which stand to benefit more from improved scoring.

Mass spectrometry related innovations

We present an improved dia-PASEF analysis pipeline which delivers high performance in identification and quantification, to an important extend thanks to the implementation of artificial intelligence technology into Spectronaut.