DIRECT SEARCHING OF DIA DATA CATCHES UP WITH SAMPLE-SPECIFIC LIBRARIES

Oliver M. Bernhardt[1], Timothy Man[2], Lucie Piecková[1], Lynn Verbeke[1], Maximilian J. Helf[1], Tejas Gandhi[1], Roland Bruderer[1], Lukas Reiter[1]

1) Biognosys AG, Wagistrasse 21, 8952 Schlieren (Zurich), Switzerland

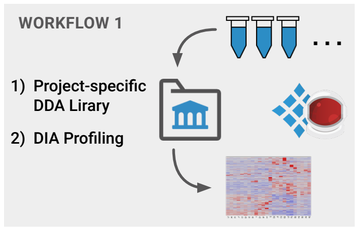
BIOGNOSYS
NEXT GENERATION PROTEOMICS

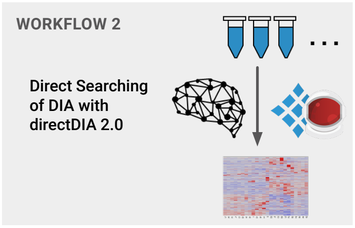
Lukas Reiter, PhD
CTO

lukas.reiter@biognosys.com
www.biognosys.com

INTRODUCTION

Label free data independent acquisition (DIA) is increasingly used for large scale proteome profiling. Typically, as a first step in this workflow, a project-specific library is generated using data dependent acquisition (DDA). This library generation step significantly complicates the DIA workflow. Alternatively, DIA data can be searched directly using a protein sequence (FASTA) file. In the past however, proteome coverage lagged behind compared to using libraries. Here, we present a deep learning enhanced algorithm, directDIA 2.0, to directly search DIA using a FASTA file. Fragmentation and retention time are predicted on the fly and used for scoring thereby significantly improving proteome coverage.



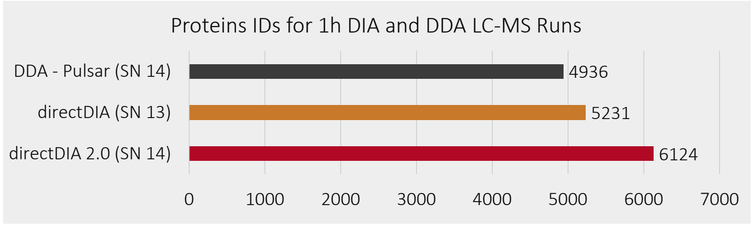


CONCLUSIONS

- **Deep learning augmented directDIA significantly improves proteome coverage**
- **To our knowledge the first example of DIA outperforming DDA significantly (25%) in protein IDs from single shot HeLa acquisitions when searching with a FASTA file**
- **directDIA 2.0 is on a par with deep project-specific DDA libraries. This largely renders library generation for DIA obsolete.**
- **directDIA 2.0 implemented in Spectronaut 14 is an easy to use and powerful quantitative proteomics workflow**

RESULTS

Figure 1: Single Shot LC-MS Protein IDs
Single shot FAIMS DDA / DIA acquisitions of a HeLa sample using a 1h gradient on a Thermo Exploris 480. All data was searched separately with a human FASTA file. Various analysis pipeline versions were used all making use of the Pulsar search engine. All SN 14 pipelines (DDA and DIA) make use of deep learning.



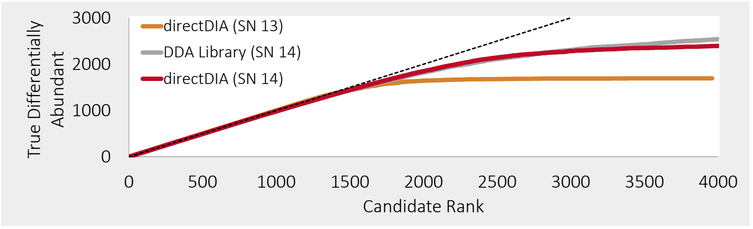


Figure 3: DirectDIA 2.0 Quantification
Comparison of the ability to detect differentially abundant proteins based on the LFQbench data set where proteomes of three organisms were mixed such that one is constant and two are varying across two conditions (Navarro et al. 2016). Data show excellent quantification with directDIA 2.0.

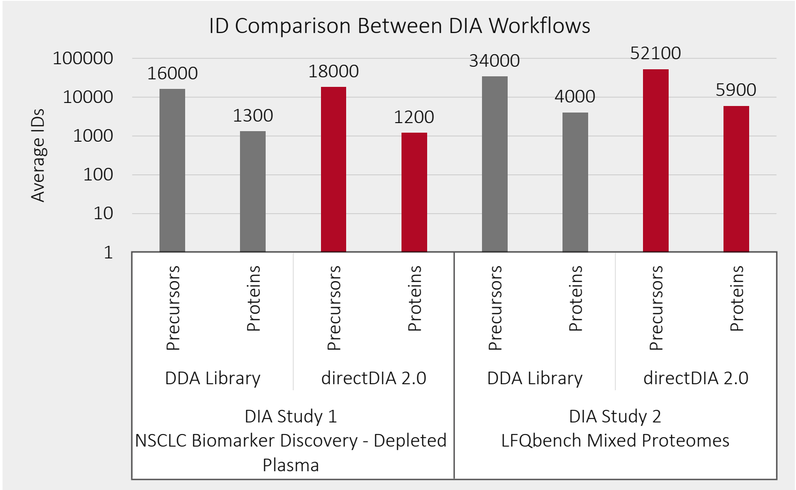


Figure 2: Comparison of DIA Workflows
Comparison of using project-specific DDA libraries and directDIA 2.0 on two different studies. Study one with human depleted plasma from healthy individuals and patients with non-small-cell lung cancer (NSCLC). And the LFQbench SWATH data set with three proteomes mixed together in controlled ratios over 2 conditions and 3 replicates (Navarro et al. 2016).

Proteome coverage achieved with directDIA 2.0 is in a similar range as when using a project-specific library. Many more projects including some sample to sample (biological) variation investigated showed the same tendency (data not shown).