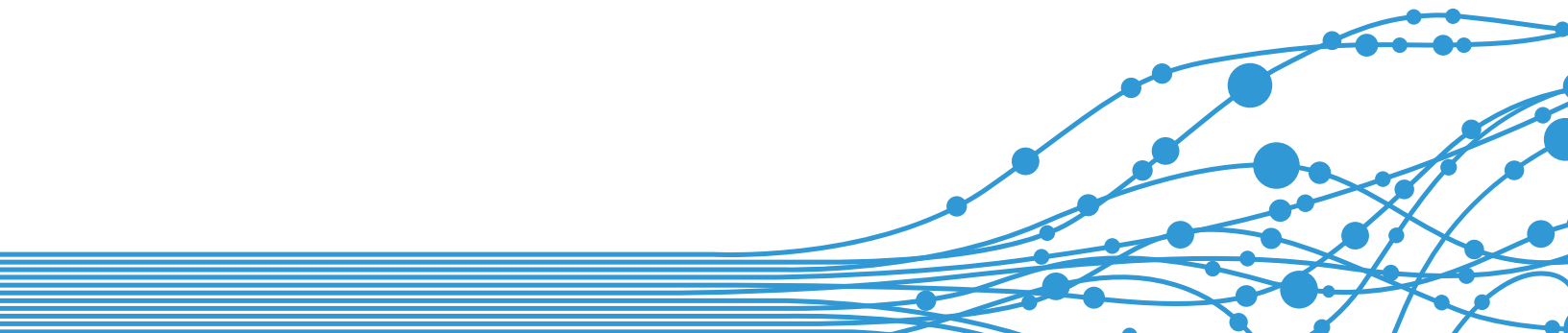
BIOGNOSYS
NEXT GENERATION PROTEOMICS

TECHNICAL NOTE

A new era in proteomics: spectral library free data independent acquisition (directDIA)

A new workflow with the simplicity of shotgun proteomics and the quantitative precision and reproducibility of data independent acquisition

If you have any questions about the technical note or Biognosys' discovery proteomics platform please contact us at support@biognosys.com.



directDIA (dDIA)
A NEW ERA IN PROTEOMICS

Introduction

Discovery proteomics aims to understand global proteome dynamics for instance in cells, tissue or an organism. The key to gain a comprehensive picture of the biology lies in the quantitative precision, reproducibility and the unbiased nature of analysis at the highest possible protein coverage.

Modern discovery proteomics workflows mainly rely on high-resolution LC-MS/MS instruments. Until today data dependent acquisition based proteomics (DDA) is the most widely used discovery proteomics technique. A major limitation of DDA based proteomics however, is the semi-stochastic peptide selection for its identification. Therefore, the same peptides will not be identified reproducibly even when analyzing technical replicates. Despite the possibility of performing MS1 alignment, this results in high number of missing values in the data matrix that lead to problems in statistical analysis and an incomplete biological picture.

Recently, data independent acquisition methods (DIA) have emerged as an alternative where in a single measurement all detectable peptides can be quantified with high sensitivity, quantitative precision and reproducibility. However, for best performance, a spectral library typically generated from DDA runs is necessary for targeted data analysis of DIA. This increases the instrument run time and especially in small experiments presents a significant cost overhead.

Biognosys, the leading developer of next-generation proteomics solutions, is now adding a new workflow called directDIA to its Spectronaut software. DirectDIA enables reproducible and precise quantification of thousands of proteins in a single sample without the need for DDA based spectral libraries. This is a simple workflow for label-free proteome quantification that offers significant savings in instrument time while maintaining high quantitative precision and high reproducibility at the same level as the targeted analysis of DIA data using spectral libraries.

In this technical note, we present an overview of the directDIA workflow and its application to a set of six human liver samples derived from healthy and adjacent cancerous tissue. In this data set, we quantified over 70,000 peptide precursors corresponding to over 5,300 protein groups. The quantitative data completeness is >90% for peptide precursors.

Methods

Samples - Biological samples consisted of human liver biopsies of poorly differentiated tumors (cancer) and adjacent healthy tissue (healthy) from three donors (in total 6 samples). Tissues were lyzed by beadmill in 8M urea buffer. Lysates were further reduced, alkylized and trypsinized using Sample Preparation Kit (Biognosys AG) according to manufacturer's protocol. Peptides were purified by C18 spin columns (MacroSpin Column, Silica C18,

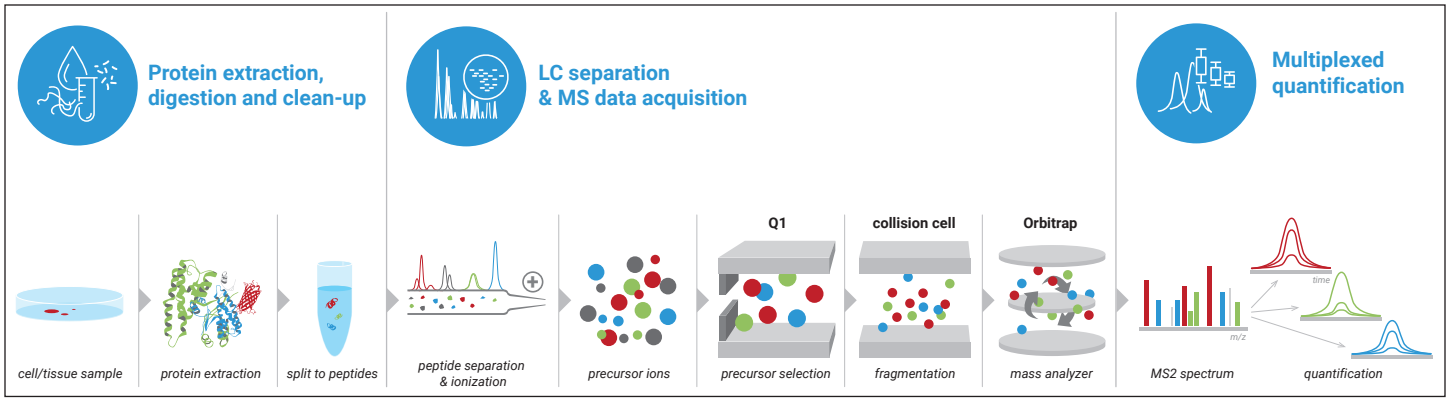


Figure 1. *Schematic representation of directDIA workflow.*

NEST group) according to manufacturer's instructions. Peptides were dried down and resuspended in loading buffer (1% acetonitrile, 0.1% formic acid in water). Peptide concentrations were determined by NanoDrop (SpectroStar Nano, BMG labtech). iRT peptides (Biognosys AG) were added according to manufactors recommendations.

Data acquisition - Two ug of the peptides mixtures were analyzed on an Easy nLC (Thermo Fisher Scientific) by a non-linear 4h gradient coupled online to a Fusion Lumos mass spectrometer (Thermo Fisher Scientific). The MS was operated in a data-independent fashion with 70 dynamic DIA segments (7 to 465 Th width dependent on the peptide precursor density) covering the mass range from 350 to 1650 Th. The resolution for the MS1 scan was set to 140k with a max injection time of 20ms and an AGC target of 5e5. The DIA scans were acquired in the Orbitrap with a resolution of 30k after fragmentation in the HCD cell (max injection time: 60ms; AGC target: 1e6; collision energy: 27%). The scan range for the DIA scans was set to 200-2000 Th.

Data analysis - The 6 DIA runs were analyzed with the new directDIA workflow in Spectronaut. For the direct database identification from the DIA runs, we used the Human Uniprot fasta file with up to 2 missed cleavages. False discovery rate (FDR) was controlled at 1 % for both PSM and protein group levels. Also, the targeted analysis was done with a 1% peptide and protein group FDR. Rest of the parameters were kept at Spectronaut Pulsar's default settings. The statistical testing for the candidates was also done in Spectronaut based on paried t-test between the two conditions corrected for multiple testing with a FDR of 5%.

Results

DirectDIA is a simple workflow for label free proteome quantification with the high quantitative precision and high reproducibility known from the targeted analysis of DIA data using spectral libraries (**Figure 1**).

In this study we acquired 6 DIA runs (3 biological replicates per condition) that were analyzed with the directDIA approach. Using the traditional spectral library based DIA workflow at least six additional DDA runs are required to build a project-specific spectral library that is used for targeted analysis of the DIA data. Thus, directDIA significantly reduced the measurement time as well as it minimized sample requirements. This presents significant savings for laboratories that are performing smaller scale discovery proteomics studies.

Using directDIA workflow, we identified and quantified in total 73'816 unique peptide precursors and 5'324 unique protein groups. Out of these, 57'379 peptide precursors and 4'935 protein groups have been quantified in all 6 samples (**Figure 2**). The quantitative data completeness for the entire data set is 93% on peptide precursor level. The identification and quantification rate were slightly decreased

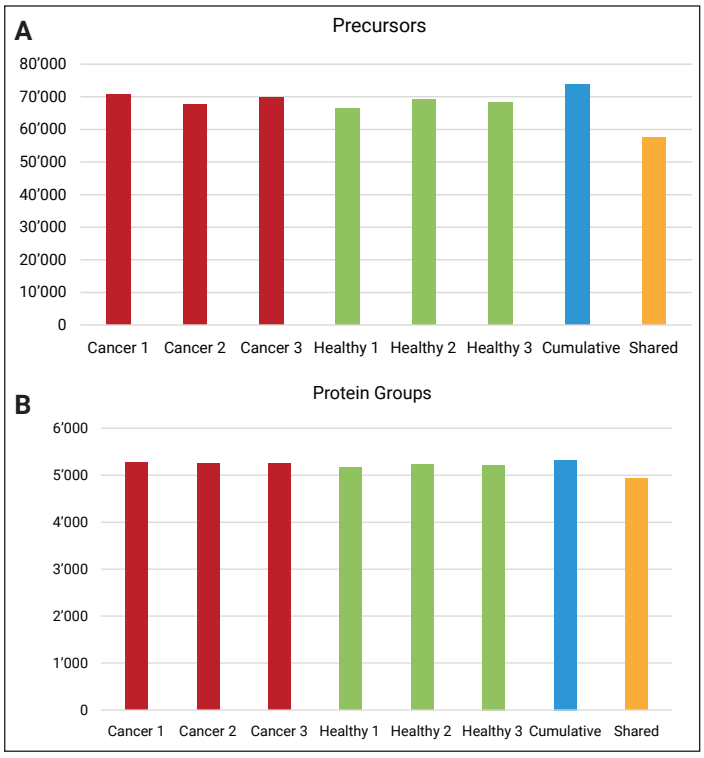


Figure 2. *Number of identified and quantified peptide precursors (A) and protein groups (B) with the directDIA workflow in cancer and healthy liver tissues. Cumulative represents the total number of unique precursors and protein groups across samples. Shared represents the number of unique precursors and protein groups identified and quantified in every sample in the dataset.*

as compared to when using a standard DIA approach with a project specific spectral library. When compared to DDA, the identification rate was similar but with a higher reproducibility and quantitative precision than typically reported from DDA experiments.

Using Spectronaut software 1892 protein groups in total were tested significantly for the differential expression profile between the cancer and the healthy tissue samples (**Figure 3**). Further analysis with fuzzy c-means clustering or pathway analysis tools can be used to determine key clusters or regulators of the proteins differentially expressed in cancer tissue. This can provide an insight in the underlying mechanisms causing cancer and identifies potential biomarkers and therapeutic targets.

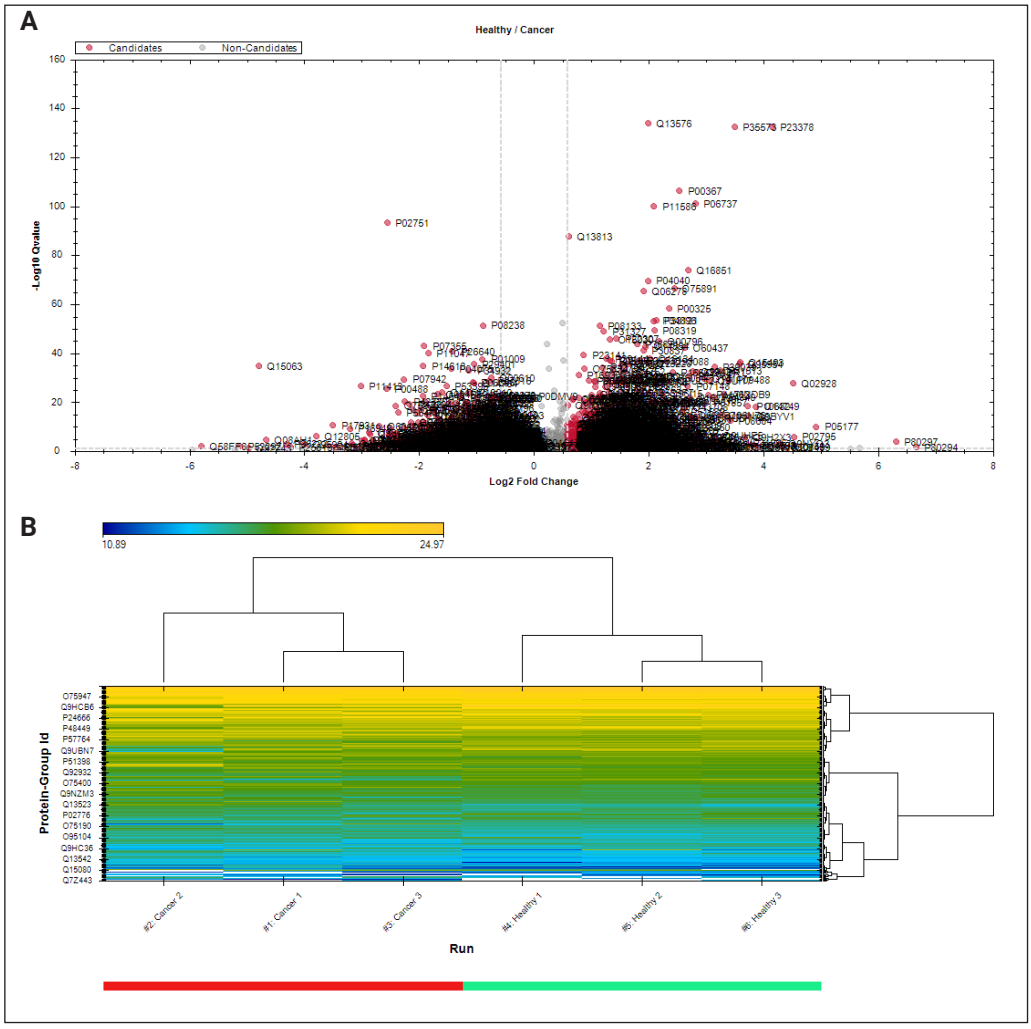
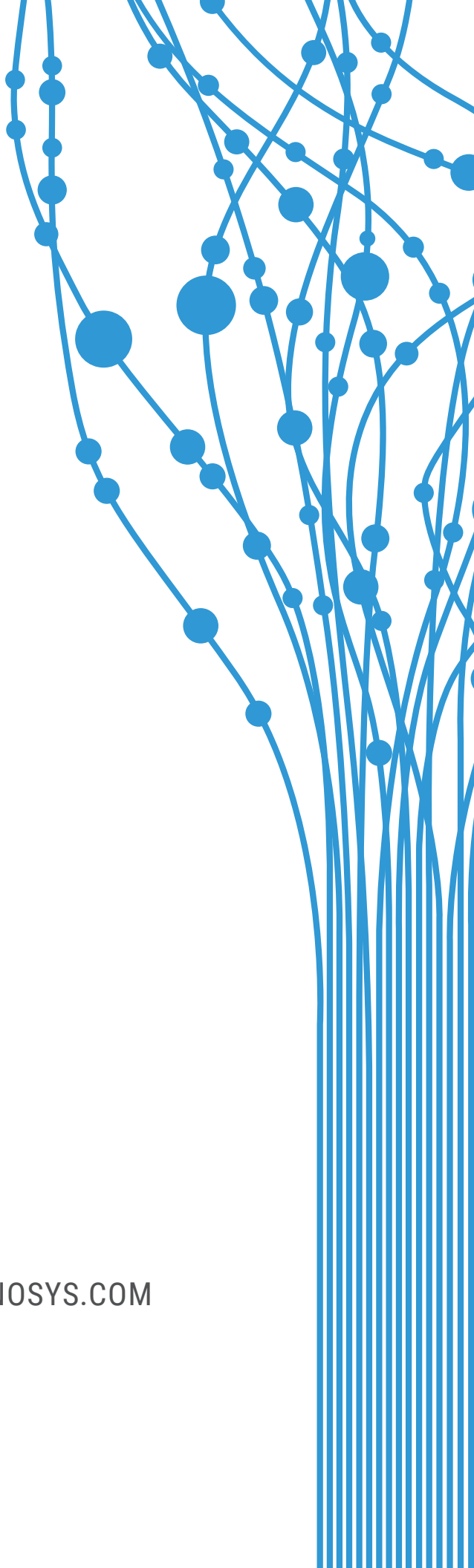


Figure 3. *Expression patterns in cancer compared to the healthy surrounding tissues shown in the volcano plot (A) and heat map (B). Data was analyzed and exported from Spectronaut software.*

Conclusions

DirectDIA presents a cost-efficient and powerful discovery proteomics approach. The results obtained are of similar high quality as standard DIA approaches where an external spectral library built from additional DDA runs is used for the targeted analysis of DIA data. Moreover, without the need for setting up the standard protocol to generate the project-specific spectral libraries directDIA workflow is much simpler to implement in the mass spectrometry based proteomics laboratory.



ABOUT BIOGNOSYS

Biognosys was founded in 2008 as a spin-off from the lab of Prof. Ruedi Aebersold at the ETH Zurich. Biognosys provides innovative services and products for protein quantification using next-generation proteomics technology.

BIOGNOSYS
NEXT GENERATION PROTEOMICS

WWW.BIOGNOSYS.COM