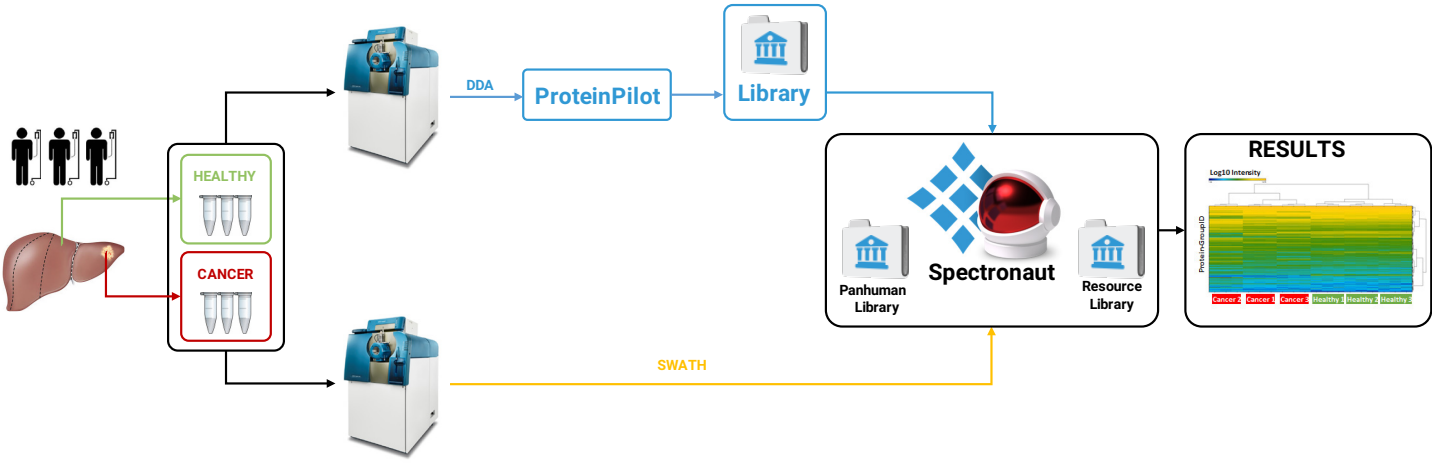
BIOGNOSYS
NEXT GENERATION PROTEOMICS

APPLICATION NOTE

FAST PROTEOME QUANTIFICATION BY MICRO FLOW SWATH® ACQUISITION AND TARGETED ANALYSIS WITH SPECTRONAUT REVEAL DEEP INSIGHTS INTO LIVER CANCER BIOLOGY



INTERESTED IN HEARING MORE ABOUT BIOGNOSYS' SOLUTIONS?

Contact us at info@biognosys.com to schedule a meeting with our experts or visit our Science Hub for additional resources we have available (papers, tutorials, webinars etc.).

INTRODUCTION

Mass spectrometry based proteomics has evolved to be the method of choice for reproducible and comprehensive proteome quantification. Currently available tools allow for the quantification of 1000's of proteins in a single run. Data-independent acquisition (DIA) methods such as SWATH® Acquisition prove to greatly improve data completeness and precision of quantification. SWATH® Acquisition is typically combined with very sensitive nanoflow LC setups; however this limits the throughput of the method because of the long time needed for column washing and re-equilibration (for example).

To improve the throughput for large scale studies, very robust microflow LC setup had to be established. With increased sample loading compared to nanoflow separations, it is possible to achieve comparable overall workflow sensitivity, making this approach a promising and robust alternative for high throughput proteomics[1].

We applied a microflow SWATH® Acquisition workflow to a small liver cancer pilot study, analyzing a set of three liver cancer samples as well as matched adjacent healthy tissue.

The 6 biological samples were acquired in triplicates. By taking advantage of micro flow the entire set of 18 samples could be measured in less than one day (1h/run). Downstream analysis was performed with Spectronaut followed by pathway analysis using Ingenuity Pathway Analysis (IPA) to reveal important insight into liver cancer biology. This workflow example can be extended to an even larger cohort study.

METHODS

Sample Preparation

The liver samples (3 cancer tissues and 3 matched adjacent healthy tissue) were lyzed by bead mill in 8M urea buffer. Lysates were prepared according to the Sample Preparation Kit Pro (Biognosys).

Peptide samples were cleaned by spin columns (MacroSpin Column, Silica C18, NEST group) according to manufacturer's instructions. Peptides were dried and re-suspended in a loading buffer including iRT peptides (Biognosys) for automated normalization retention time and mass calibration with Spectronaut.

Chromatography

Separation of the trypsin digestion of the human liver tissue extract was performed with a NanoLC™ 425 System (SCIEX) operating in a trap elute injection mode at micro flow rates.

A combination of 0.3x150 cm analytical column and a 0.5 x 5mm trap column both packed with Triart C_{18} 3um (YMC Dinslaken) was used.

Samples were loaded for 3min at 10µl/min into the trap column equilibrated in 2% acetonitrile/98% water/0.1% formic acid and then eluted with a short gradient (4-32% solvent B in 43 min, (B: 0.1% formic acid in acetonitrile) at 5µL/min (total run time 57min). The total peptide amount injected on column was 8µg.

Mass Spectrometry

The MS analysis was performed with the TripleTOF® 6600 system (SCIEX) using the Turbo V™ Source with 50µm I.D. hybrid electrodes (SCIEX).

Data dependent acquisition was acquired on the unfractionated lysates, using the 30 MS/MS per cycle with a 50 ms accumulation time. The variable window SWATH® Acquisition method was built using Analyst® TF Software 1.7.1 (SCIEX) with 100 variable windows and 25 ms accumulation times, covering a precursor mass range from 400 to 1250 m/z.

Data Processing

For the spectral library generation, first the data dependent acquisition (DDA) files were processed using ProteinPilot™ 5.0 Software (SCIEX) and filtered at a 1% global FDR.

Search results were utilized with Spectronaut (Biognosys) to generate the spectral library for the targeted data extraction of the SWATH® Acquisition data using the default parameters (DDA library: 22,506 precursors; 18,253 peptides; 2,570 protein groups).

We also analyzed the data with the Pan Human Library[2] (206,700 precursors; 147,526 peptides; 10,596 protein groups) and the publicly available liver cancer library with Spectronaut (82,821 precursors; 64,325 peptides; 7,188 protein groups).

The raw data files were directly processed with Spectronaut and the results filtered by a FDR of 1% on the precursor and protein level.

The quantitative data was filtered by a sparse percentile = 0.5, meaning that the precursors need to be assuredly identified (Q value < 0.01) in at least 50% of samples to be considered for the downstream analysis.

Statistical testing of the data was also performed with Spectronaut (Student t-test, Q value < 0.01, fold-changes > 2 folds). Biological interpretation of the data was done by IPA (QIAGEN Bioinformatics).

RESULTS

The combination of microflow LC separations and SWATH® acquisition enabled the acquisition of a sample set of 18 liver cancer samples including run triplicates in less than 1 day (~1h per sample). The data was then analyzed by Spectronaut, which performed an automated mass calibration on MS1 and MS2 level as well as the retention time dimension (based on the iRT concept[3,4]).

Spectronaut also performed a protein grouping based on the IDpicker5 algorithm by using the peptide identifications and the fasta database provided by the user. By application of the DDA library, almost the whole library was recovered in the data (22,477 precursors; >99% and 2,560 protein groups, >99%). To improve the certainty in the data, we filtered the data by using the sparse percentile filter (set to 0.5) with Spectronaut.

The filter ensured that in downstream statistical testing, only precursors would be considered that were identified in at least 9 out of the 18 runs. The rationale behind this approach was to include precursors that might be absent in the either the cancer or the healthy cohort.

With this filter 2,362 protein groups (92% recovery) were quantified (19,983 precursors, 89% recovery). On average roughly 10% more protein groups were identified in the cancer cohort compared to the healthy cohort (2,404 vs 2,199), indicating large alterations in the proteome composition in cancer tissue samples.

Spectronaut performed automatically an unsupervised clustering analysis and displayed the results in a heat map (***Figure 1A***).

This clustering revealed that the run replicates the cluster closely together, indicating that technical replicates can be omitted to increase the number of biological samples (CVs < 10%). The cancer and healthy cohort were clearly separated. Protein expression levels across samples within the cancer and healthy cohort were calculated to assess the biological variance of each protein.

In healthy tissues the proteome composition was tightly regulated, indicated by a consistent protein expression across biological samples. Cancer tissues on the other hand were marked by a higher heterogeneity as visualized in the heat map. To identify the differentially expressed protein groups, a pairwise Students t-test was applied.

In total 594 protein groups were found to differ in abundance between the healthy and cancer cohort (Q < 0.01 with minimum of 2-fold change, visualized as volcano plot in ***Figure 1B***).

Of these proteins 18 were reported as hepatocellular markers (according to oncomine.org). The results were loaded into a IPA to perform a biological interpretation of the data. The pathway analysis in the IPA showed that the enrichment of several degradation pathways (***Figure 2A***).

Most proteins in these pathways were less abundant in the cancerous tissue indicating an impaired detoxification capability. We also analyzed the upstream regulators and the micro RNA mir-122, a known tumor suppressor with potential antitumor function6 was the most significantly regulated molecule (p = 3.1e-10, activation z-score = -3.133).

Figure 1: Data Analysis Overview

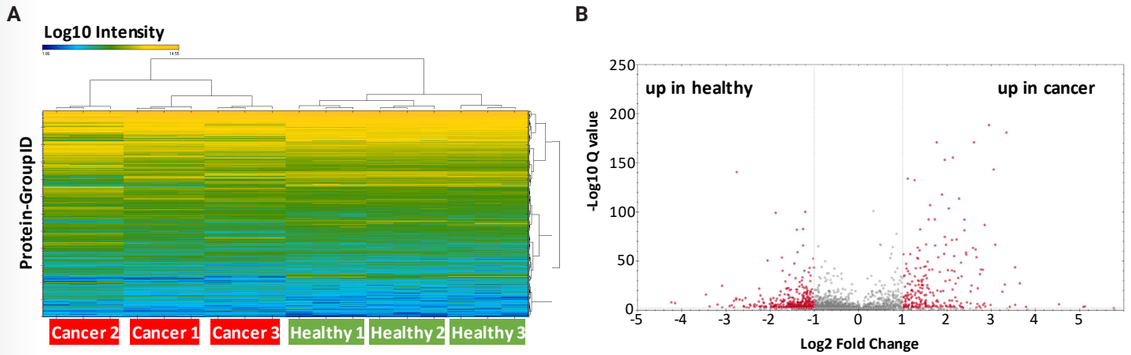


Figure 1: A) Heat map showing the unsupervised clustering of the samples based on protein intensities. Each biological sample consists of 3 run replicates which closely clustered together (not labeled separately). B) Volcano plot visualizing the protein fold changes and associated Q values based on Student t-test. Red dots indicate differentially abundant proteins (Q value < 0.01, fold change > 2 folds).

Figure 2: Deriving Biological Meaning from the Protein Expression Differences

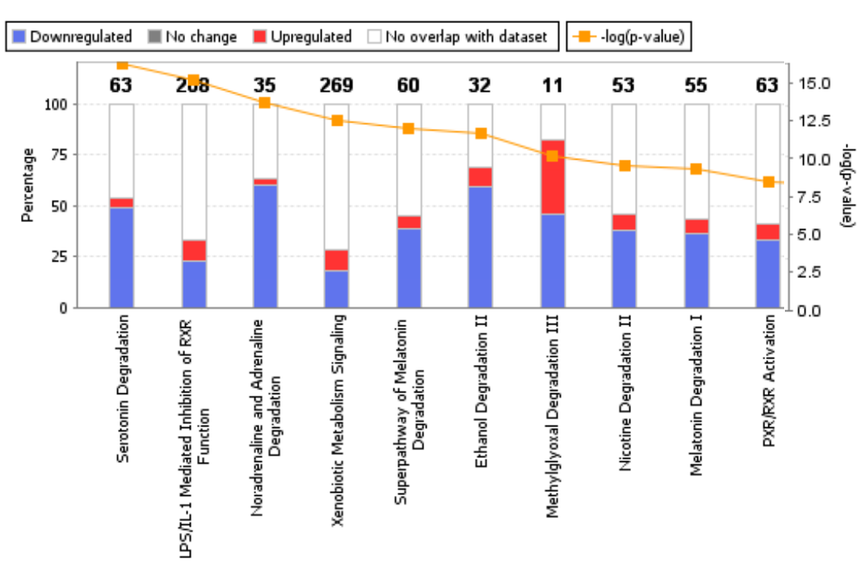


Figure 2. A) Pathway analysis of significantly regulated proteins.

Most of the proteins (15/17) within the mir-122 network showed the predicted regulation (***Figure 2B***). A more detailed expression analysis showed that the expression changes of some proteins were consistent in the cancer cohort, e.g. Septin-9 (SEPT9), quantified at comparable elevated levels in all cancer samples (***Figure 2C***).

Other proteins, e.g. Cytochrome P450 2C9 (CP2C9), had distant differences in the expression level in one of the three cancer samples (***Figure 2C***). Although we only analyzed a small cohort, the examples demonstrate how the data helped to gain meaningful insight into cancer biology and cancer heterogeneity.

Figure 2: Deriving Biological Meaning from the Protein Expression Differences

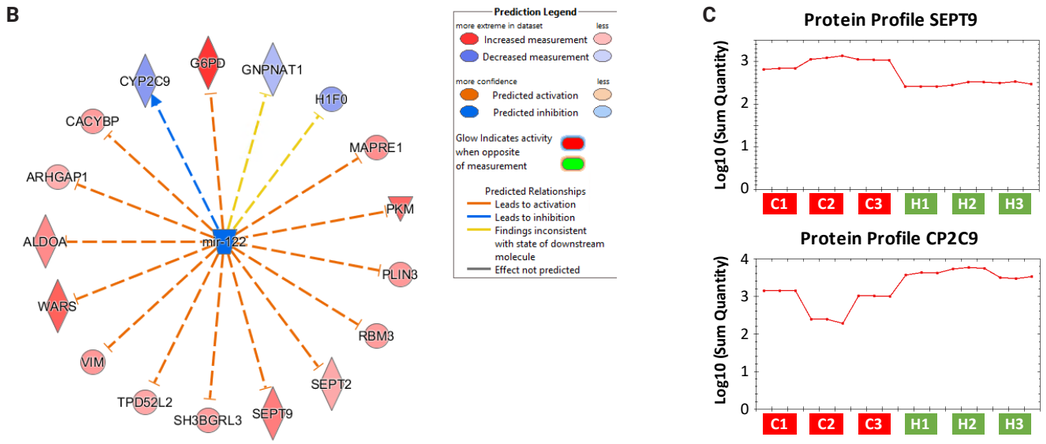


Figure 2. B) Network of tumor suppressor mir-122. C) Examples of regulated proteins in the mir-122 network.

A short investigation of the impact of library size on results was also undertaken. We analyzed the data with the DDA library created on this exact sample set as well as two publicly available libraries; a liver cancer library and the pan human library[2] (both available with Spectronaut). The recovery of the currently used DDA library was 92% (protein group level).

This implies that the size of the library limits the analysis in terms of quantified proteins, therefore application of a larger (resource) library might increase the number of quantified proteins.

Both of the publicly available libraries contain more than 80,000 precursors and more than 7,000 protein groups potentially enabling a high analytical depth for our small pilot study.

The application of both resource libraries greatly improved the number of quantified precursors, peptides and protein groups (***Figure 3A***, liver library: 43,501 precursors (+94%), 35,475 peptides (+94%), 3,713 protein groups (+45%); pan human library: 35,791precursors (+59%), 29,481 peptides (+62%), 3,120 protein groups (+22%)).

However, by application of the previously applied percentile filter these numbers dropped a little below the number of quantifications from the data analysis by the DDA based library. We also compared the overlap in quantified protein groups between the datasets from the different libraries.

To minimize the influence of differential protein grouping between the three datasets, only the first entries of the protein groups were compared (***Figure 3B***).

The majority of all proteins were quantified in all three datasets (1,619; 60.8%) and only a small proportion in only one dataset (DDA based library: 342, 12.8%; liver library: 204, 7.7%; pan human library: 80. 3%). This result clearly indicates that resource libraries can also be used in the data analysis of SWATH® acquisition enabling to omit the time-consuming spectral library generation step.

Summary

Within less than 1 day, a small study of liver cancer proteomics was performed. Because of the throughput and robustness of the microflow setup the number of analyzed samples can be easily increased. More of the sample is required to run the microflow SWATH® acquisition workflow (~4x more)[1], however this amount is often readily obtained from biological studies. With this robust sample acquisition strategy combined with powerful software to analyze the data, large sample cohorts can now be analyzed within a reasonable time frame (~150 proteomes per week).

Figure 3: Application of Resource Spectral Libraries

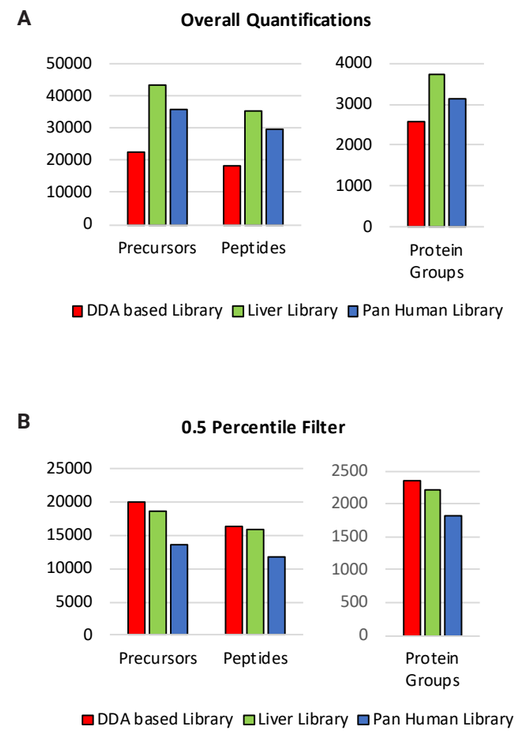


Figure 3. A) Left panel shows the number of quantified precursors, peptides and protein groups in the microflow SWATH runs by application of three different libraries (red: DDA based library generated in the same study, green: deeper liver spectral library generated on the same type of sample, blue: pan human library). Right panel shows the same data after application of the 0.5 percentile filter in Spectronaut. B) The Venn diagram provides an overview of the quantified protein groups by application of three previously mentioned libraries (red: DDA based library, green: liver library, blue: pan human library).

ACKNOWLEGEMENTS

SCIEX for acquisition of the data

1. Microflow SWATH® Acquisition for Industrialized Quantitative Proteomics, SCIEX technical note, RUO-MKT-02-3637-A.

2. Rosenberger, G. et al. A repository of assays to quantify 10,000 human proteins by SWATH-MS. Sci. Data 1, 140031 (2014).

3. Escher, C. et al. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics 12, 1111–1121 (2012).

4. Bruderer, R., Bernhardt, O., Gandhi, T. & Reiter, L. High precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. Proteomics 1–20 (2016). doi:10.1002/pmic.201500488.This

5. Zhang, B., Chambers, M. C. & Tabb, D. L. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. J. Proteome Res. 6, 3549–57 (2007).

6. Nakao, K., Miyaaki, H. & Ichikawa, T. Antitumor function of microRNA-122 against hepatocellular carcinoma. J. Gastroenterol. 49, 589–593 (2014).

7. https://biognosys.com/media/2016large-scale-dia-data-analysis-and-scalability-using-spectronaut-software.pdf