A Data Independent Acquisition Workflow Enables the Profiling of Thousands of Human Cancer Tissues for Precision Oncology

Jakob Vowinckel[1], Karel Novy[1], Thomas Corwin[2], Tobias Treiber[1], Roland Bruderer[1], Lukas Reiter[1], Eike-Christin von Leitner[2], Oliver Rinner[1], Claudia Escher[1]

1) Biognosys AG, Schlieren, Switzerland, 2) Indivumed GmbH, Hamburg, Germany

INTRODUCTION

Precision oncology requires deep understanding of molecular mechanisms involved in cancer biology. The combined analysis of several types of omics data will generate knowledge that goes beyond purely genetic approaches on which precision medicine has relied on almost exclusively in the past. With the rise of data-independent acquisition (DIA) mass spectrometry and advances in chromatography, proteomics technology has come into close proximity of the scale and depth of next-generation sequencing required for multi-omics efforts.

Here we present a platform for the acquisition of true large-scale proteomics and phospho-proteomics datasets for a unique collection of bio-specimens derived from the IndivuType cohort of Indivumed, Germany. Matching fresh frozen tumor and adjacent normal tissue samples from thousands of patients (Figure 1) were obtained from Indivumed's global network of partner hospitals. These samples are of outstanding high quality as they are collected following strictly defined standard operating procedures minimizing variation from pre-analytical variables.

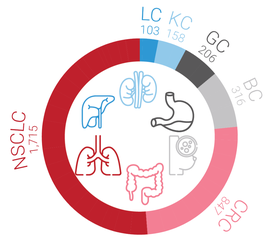

Figure 1: Overview of tumor types included in this study.

Out of in total 3,345 samples measured, 1,715 samples were from non-small cell lung carcinoma (NSCLC) patients, 847 samples from colorectal cancer (CRC) patients, 316 samples from breast cancer (BC) patients, 206 samples from gastric cancer (GC) patients, 158 samples from kidney cancer (KC) patients, and 103 samples from liver cancer (LC) patients.

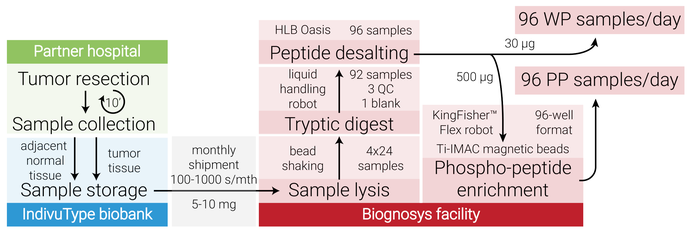

Figure 2: Specimen collection and sample preparation workflow diagram.

Tumor resection is carried out in partner hospitals in Europe, US, and Asia under unified standard operating procedures (SOP). After maximal 10 min following resection samples from tumor and adjacent tissue are frozen and transferred to IndivuType biobank. Monthly shipments to Biognosys facility typically comprise several 100s of samples per month, and sample preparation is performed in batches of 92 samples. After lysis of 5-10 mg of tissue using bead shaking, reduction, alkylation and tryptic digest is performed in 96-well format using a liquid handling robot. With every batch of 92 tissue samples 3 quality control (QC) samples and one blank sample are processed. After peptide desalting using Waters HLB Oasis 96-well plates, 500 µg of peptide material is subjected to phospho-peptide enrichment and 30 µg is used for whole-proteome (WP) analysis. Phospho-peptide enrichment is carried out with the help of a KingFisher™ Flex robot using MagReSyn™ Ti-IMAC magnetic beads in 96-well format. This pipeline allows preparation of 96 WP and 96 phospho-proteome (PP) samples per day.

METHODS

Sample processing (Figure 2) was performed by first placing 5-10 mg of fresh-frozen tissue in 2 mL Precellys® CK14 tubes containing 1.4 mm ceramic beads. For each batch of 92 samples, 4 batches of 20 or 24 samples were disrupted by adding lysis buffer containing PhosSTOP™ and bead shaking using a Precellys® Evolution Homogenizer equipped with a Cryolys® cooling module. After centrifugation, supernatant was transferred to 96-deep well plates. 3 workflow quality control (QC) samples were added to each plate from a previously generated lysate from healthy liver, lung, and colon tissue with one blank sample. Lysate protein concentration was determined using BCA assay, and proteins were reduced and alkylated. With the help of a liquid handling robot, samples were diluted and trypsin added. After overnight digest samples were acidified and peptides desalted using Waters HLB Oasis 30 mg 96-well plates. 500 µg of peptide preparation was subjected to phospho-peptide enrichment using MagReSyn® Ti-IMAC magnetic beads (ReSyn Biosciences) essentially as described [1] with modifications to enable processing using a KingFisher™ Flex robot equipped with a 96-magnetic pin head. Peptides were desalted using Waters µElution plates, dried down and resolubilized.

For DIA LC-MS/MS measurements, 5 µg of peptides per sample were injected to a reversed phase column (nanoEase M/Z Peptide CSH C18 Column, 1.7 µm, 300 µm X 150 mm) on a Waters ACQUITY UPLC M-Class LC connected to a Thermo Scientific™ Orbitrap Q Exactive™ HF-X mass spectrometer equipped with an EASY-spray source. The nonlinear LC gradient was 1 - 60 % solvent B in 45 minutes (WP) or 60 min (PP) at 50°C and a flow rate of 5 µL/min. The DIA method consisting of one full range MS1 scan and 50 DIA segments was adapted from Bruderer *et al.* [2].

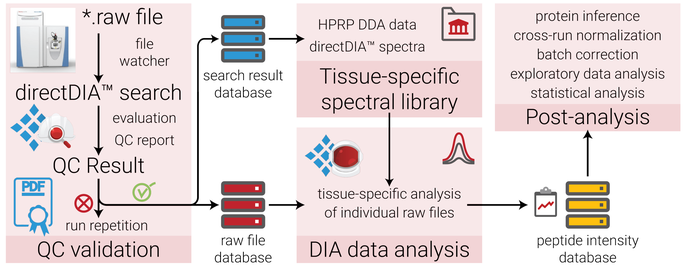

Figure 3: Quality control and DIA data analysis workflow

Automated QC validation and DIA data analysis is performed by a series of scripts that first copy new raw files to a workstation and initiate directDIA™ search using SpectroMine™ software. Results are compared to pre-defined thresholds, and a QC certificate is generated. Runs passing thresholds are added to a raw file database, and corresponding directDIA™ spectra to a search results database. Together with DDA data of HPRP fractionated sample pools the latter is used to generate a comprehensive hybrid spectral library for each tumor type. DIA data analysis of individual raw files is performed using Spectronaut™ and the matching tissue-specific spectral library, and peptide intensity data is added to a peptide intensity database. This growing body of data is then used to perform post-analysis.

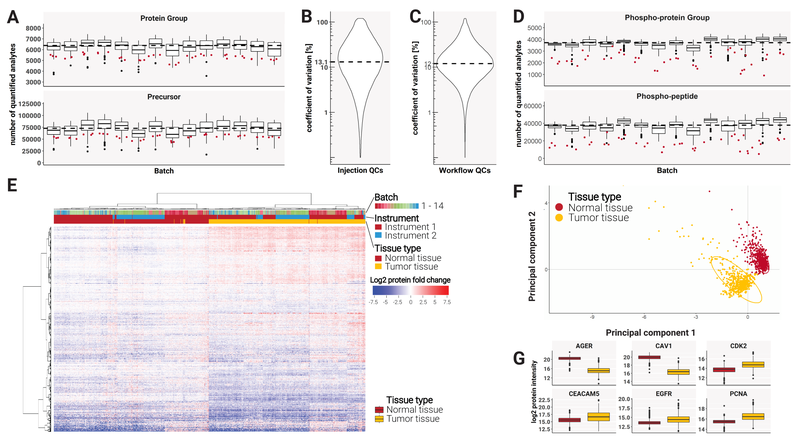

Figure 4: Data obtained is highly reproducible in terms of number of quantified analytes, technical variation and phospho-peptide enrichment.

(A) The average number of protein groups and precursors quantified with peak-picking q-value < 0.01 is shown for 1,192 samples in 14 batches analyzed using WP workflow. Red dots indicate workflow QC samples. **(B)** WP protein-level coefficients of variation (CVs) for 56 injection QC samples acquired over 28 batches are on median 13.1 %. **(C)** WP protein-level CVs for 42 workflow QC samples acquired over 14 batches are on median 12.0 %. **(D)** The average number of phospho-protein groups and phospho-peptides quantified with q-value < 0.01 is shown for 1,188 samples in 14 batches analyzed using PP workflow. Red dots indicate workflow QC samples. **(E)** WP protein-level heat map of 998 NSCLC samples shows clustering according to tissue type, and minor clustering according to instrument used. Normal tissue samples show lower abundance for a large number of proteins. **(F)** WP principal component analysis shows co-clustering of 488 normal tissue samples and 510 tumor tissue samples in components 1 and 2. **(G)** Protein intensity of six known markers for NSCLC tumors are shown confirming robust protein change in tumor compared to normal tissue.

Data validation (Figure 3) was performed using an automated pipeline consisting of a series of R scripts orchestrated with the help of an R/shiny app. After LC-MS/MS acquisition files were copied to a workstation and subjected to directDIA analysis using SpectroMine™ software (Biognosys). The resulting identifications together with TIC intensity and other parameters were compared to pre-set thresholds. A QC certificate was issued, and passing runs were moved to a raw file repository. Spectral libraries were generated separately for each tissue by combining data-dependent acquisition (DDA) data from 15 high-pH reversed phase (HPRP) fractions of 100 pooled samples with all directDIA spectra of the given tissue type. Interference correction, peptide selection and protein inference were trained using machine learning in Spectronaut™ software (Biognosys) and applied to all analyzed files. Automated Spectronaut™ analysis using trained tissue-specific spectral libraries was carried out individually for each raw file, and export tables were stored in a repository. For global analysis data for the desired samples was read and subjected to normalization and post-analysis.

RESULTS AND CONCLUSIONS

- The optimized platform for large-scale analysis of the tumor proteome allows processing and analysis of 976 WP or 732 PP samples per month, and multiplexing using several instruments is possible
- WP analysis has an average depth of 6,380 protein groups (72,772 precursors), and PP analysis an average depth of 37,926 phospho-peptides (3,712 phospho-protein groups)
- WP median coefficients of variation are 13.1 % for injection QC samples and 12.0 % for workflow QC samples
- The automated pipeline setup allows near real-time validation of LC-MS/MS performance, and direct access to peptide and protein level intensity data without long processing times
- Data is continuously generated and integrated into Indivumed's IndivuType multi-omics database, supporting identification and validation of new molecular cancer drug targets and biomarkers
- Advances in micro-flow LC and DIA MS together with powerful computational processing make proteomics accessible to precision oncology approaches

REFERENCES

[1] Bekker-Jensen, D.B. et al. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. Nat Commun 11, 787 (2020)

[2] Bruderer et al., Analysis of 1508 plasma samples by capillary flow data-independent acquisition profiles proteomics of weight loss and maintenance. MCP 18 (6), 1242-1254 (2019)

CONTACT INFORMATION

Jakob Vowinckel, PhD
Senior Scientific Project Manager

Phone +41 44 738 20 41
jakob.vowinckel@biognosys.com

Biognosys AG
Wagistrasse 21
8952 Schlieren
Switzerland

www.biognosys.com

SCAN QR CODE TO REQUEST POSTER!