



OPEN

DATA DESCRIPTOR

A hybrid DDA/DIA-PASEF based assay library for a deep proteotyping of triple-negative breast cancer

Petr Lapcik¹, Klara Synkova¹, Lucia Janacova¹, Pavla Bouchalova¹, David Potesil², Rudolf Nenutil³ & Pavel Bouchal¹  

Triple-negative breast cancer (TNBC) is the most aggressive subtype of breast cancer, and deeper proteome coverage is needed for its molecular characterization. We present comprehensive library of targeted mass spectrometry assays specific for TNBC and demonstrate its applicability. Proteins were extracted from 105 TNBC tissues and digested. Aliquots were pooled, fractionated using hydrophilic chromatography and analyzed by LC-MS/MS in data-dependent acquisition (DDA) parallel accumulation-serial fragmentation (PASEF) mode on timsTOF Pro LC-MS system. 16 individual lysates were analyzed in data-independent acquisition (DIA)-PASEF mode. Hybrid library was generated in Spectronaut software and covers 244,464 precursors, 168,006 peptides and 11,564 protein groups (FDR = 1%). Application of our library for pilot quantitative analysis of 16 tissues increased identification numbers in Spectronaut 18.5 and DIA-NN 1.8.1 software compared to library-free setting, with Spectronaut achieving the best results represented by 190,310 precursors, 140,566 peptides, and 10,463 protein groups. In conclusion, we introduce assay library that offers the deepest coverage of TNBC proteome to date. The TNBC library is available via PRIDE repository (PXD047793).

Background & Summary

Triple-negative breast cancer (TNBC) is an aggressive disease that represents about 15% of all breast cancer cases and is typically associated with a poor prognosis^{1,2}. Approximately 28–39% of the TNBC patients do not survive 5 years after the diagnosis, leading to shorter survival rates compared to other breast tumor subtypes³. TNBC tumors are defined by the lack of expression of estrogen receptor, progesterone receptor and human epidermal growth factor receptor-2 (HER2)⁴. The absence of these receptors that are targets of biological therapy for other breast cancer subtypes results in insensitivity of TNBC tumors to hormonal and anti-HER2 therapies, leading to untargeted cytotoxic chemotherapy as a mainstay systemic treatment of TNBC tumors^{5,6}. Moreover, TNBC tumors exhibit a significant molecular heterogeneity⁷ that complicates development of targeted therapies for TNBC patients. A deep proteome coverage of TNBC tumors is essential for development of precise classification of TNBC tumors and identification of novel therapeutic targets.

Mass spectrometry-based techniques allow large-scale identification and quantification of proteins and show potential for addressing clinical issues⁸. The data-independent acquisition (DIA) strategies offer accurate and reproducible protein quantification across large groups of samples with reduced missing value numbers^{9,10}, as demonstrated also for breast cancer tissues¹¹. To assign the elution groups of fragment ions to precursors and peptides, analysis of DIA data requires a spectral library. Typically, a library-based approach is being used, and the spectral library can be generated from data-dependent acquisition (DDA) LC-MS/MS measurements of individual samples or their fractionated pools^{12,13}. As an alternative, an approach that implements libraries predicted *in silico* from protein sequence databases or constructed from DIA data as pseudospectra have been employed^{14,15}.

¹Department of Biochemistry, Faculty of Science, Masaryk University, Brno, Czech Republic. ²Central European Institute of Technology, Masaryk University, Brno, Czech Republic. ³Department of Oncological Pathology, Masaryk Memorial Cancer Institute, Brno, Czech Republic. ✉e-mail: bouchal@chemi.muni.cz

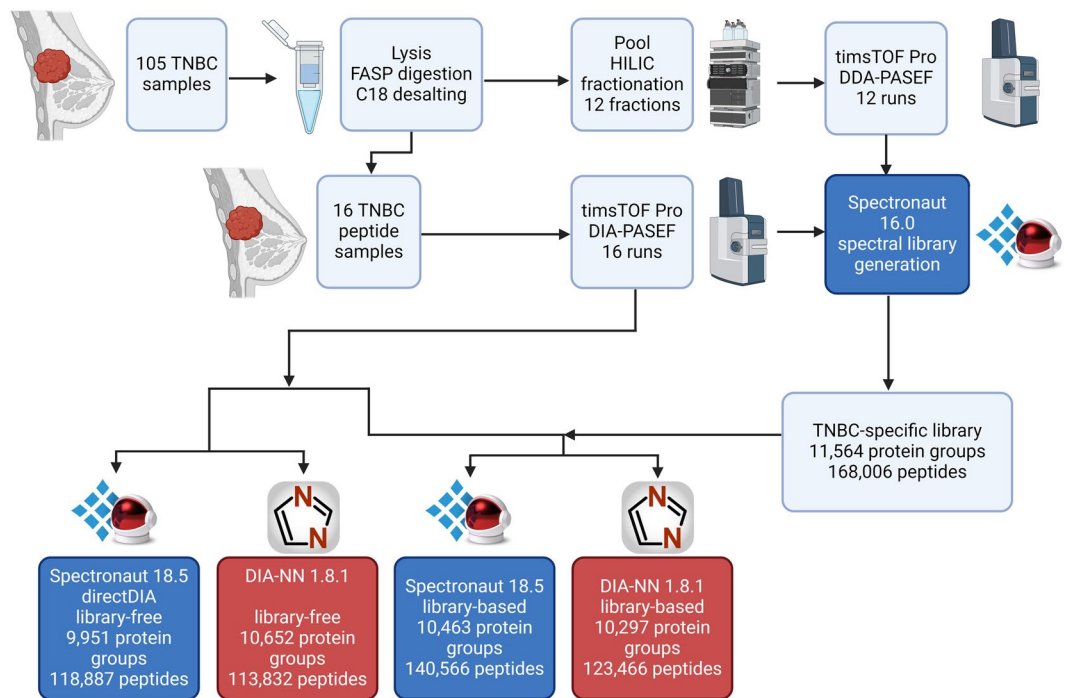


Fig. 1 Workflow of the study. Generation of TNBC-specific targeted proteomics assay library using data from fractions of the pool of 105 TNBC tissues and 16 individual samples and quantitative data extraction in Spectronaut and DIA-NN in library-based and library-free settings. The figure was created with BioRender.com.

Furthermore, the use of sample type-specific spectral libraries is beneficial for precise peptide and protein identification¹⁶ and can offer better proteome coverage compared to pan-human libraries¹⁷. García-Adrián *et al.* recently published a DIA dataset including 125 TNBC tumors measured on LTQ-Orbitrap Fusion Lumos mass spectrometer that covers 3092 proteins, of which 1206 proteins were consistently detected and quantified in at least 66% of the samples¹⁸. The application of this dataset for definition of TNBC subgroups and novel therapeutic targets is however limited by the low number of protein identifications. As novel mass spectrometry instrumentation and strategies have been developed, there is potential for generation of more comprehensive TNBC-specific DIA datasets and spectral libraries. Particularly four-dimensional proteomics technology implementing trapped ion mobility spectrometry significantly improved sensitivity of peptide detection¹⁹. This technology implemented in parallel accumulation-serial fragmentation (PASEF) strategy allows multiplication of the sequencing speed without any loss of sensitivity²⁰. PASEF can be performed in a DIA mode (DIA-PASEF) that utilizes smaller number of ions in particular DIA windows and improves specificity, offers near 100% ion utilization and high degree of reproducibility as well as quantitative accuracy²¹. Moreover, several software strategies have been developed for DIA data extraction from the raw runs²², from which, Spectronaut²³ and DIA-NN²⁴ represent two popular tools that are also capable of processing the DIA-PASEF data^{25,26}. Spectronaut implements database search engine Pulsar that allows targeted DIA data extraction using spectral libraries generated from DDA runs. Moreover, Spectronaut incorporates library-free search strategy called directDIA for targeted data extraction after a direct search of DIA data by generating pseudo MS/MS spectra¹⁴. Combination of these approaches allows generation of hybrid libraries²⁷ that combine searches of both DDA and DIA data using Pulsar search engine, resulting in a deep proteome coverage and improved retention time calibration. On the other hand, DIA-NN generates libraries by prediction from protein sequence databases for library-free searches, nevertheless it allows targeted DIA data extraction using spectral libraries generated in external tools. DIA-NN software offers a fast performance which is valuable for high-throughput applications²⁴.

To generate a comprehensive proteomics assay library specific for TNBC, we fractionated peptides from a large set of TNBC tissues, analyzed the fractions using timsTOF Pro LC-MS system in DDA-PASEF mode and 16 of the individual samples in DIA-PASEF mode, generated a new hybrid library of targeted proteomics assays in Spectronaut software, and compared the library-based and library-free DIA data extraction strategies using two key software tools, Spectronaut and DIA-NN (Fig. 1).

Methods

Sample information and patient characteristics. Aliquots of primary tumor tissue lysates from 105 patients treated at Masaryk Memorial Cancer Institute (MMCI) in Brno, Czech Republic, were used to generate the spectral library. Only samples from women diagnosed with a breast cancer with negative status of estrogen, progesterone and human epidermal growth factor receptors using immunohistochemistry, who were treated at MMCI between 2001–2015 and did not receive any neoadjuvant treatment before the tumor resection, were included into the study. The study was approved by Ethics committees of MMCI (2019/1543/MOU, 2021/1723/MOU)

and Masaryk University (EKV-2021-083). All patients have signed the informed consent. For the patient characteristics please see Supplementary File 1.

Sample preparation. The tissue samples were received in 20 min after the surgical removal, evaluated by a pathologist and snap frozen in liquid nitrogen. The pieces of tumor tissue corresponding to $2 \times 2 \times 2$ mm blocks were subjected to lysis in 200 μ L of 6 M guanidine hydrochloride and 1% Triton X-100 in 0.1 M phosphate buffer pH 6.6. The tissue samples were subsequently homogenized in the homogenizer (Retsch, Haan, Germany) for 2×2 min at 25 s^{-1} . The samples were sonicated 30×0.1 s, power 50 W with a needle sonicator (Bandelin HD 2200; Bandelin, Berlin, Germany). The protein lysates were next incubated for 60 min at RT and centrifuged at 14,000 g/20 min/4 °C. The protein concentrations in the supernatants were measured using the RC-DC protein assay (Bio-Rad, Hercules, CA, USA). The protein samples were subjected to trypsin digestion with the use of Filter-Aided Sample Preparation (FASP) method²⁸ with modifications. 100 μ g of protein per sample was transferred to the Microcon filter device, cut-off 30 kDa (Millipore) with 100 μ L of 8 M urea in 0.5 triethylammonium bicarbonate (TEAB) pH 8.5 and centrifuged at 14,000 g/15 min. The samples were reduced with 10 μ L of 50 mM tris(2-carboxyethyl)phosphine (Sigma-Aldrich) in 100 μ L of 8 M urea in 0.5 M TEAB pH 8.5 for 60 min/600 rpm and centrifuged at 14,000 g/15 min. Next, the samples were alkylated using 5 μ L of 200 mM S-methyl methanethio-sulfonate (Sigma-Aldrich) in 100 μ L of 8 M urea in 0.5 M TEAB pH 8.5 for 1 min/600 rpm and for 20 min in the dark without shaking, and centrifuged at 14,000 g/15 min. The protein samples on the filter were washed twice with 100 μ L of 0.5 M TEAB pH 8.5 and centrifuged at 14,000 g/20 min. The proteins were digested by 3.33 μ L of trypsin (Promega, 1 μ g/ μ L) in the trypsin:protein ratio 1:30 in 100 μ L of 0.5 M TEAB pH 8.5 for 16 hrs at 37 °C. Resulting peptides were eluted with centrifugation at 14,000 g/15 min and desalted on MicroSpin columns C18 (Nest Group) as described in²⁹ with modifications. First, the columns were washed twice with 200 μ L of 0.1% TFA in 100% acetonitrile (ACN) and centrifuged at 100 g/3 min. The columns were washed with 200 μ L of 0.1% TFA in water and centrifuged at 300 g/3 min. Another 200 μ L of 0.1% TFA in water was added to the columns followed by 15 min incubation and centrifugation at 300 g/3 min. The peptide samples were then added to the hydrated columns, centrifuged at 500 g/3 min, and washed three times with addition of 200 μ L of 0.1% TFA in water and centrifugation at 500 g/3 min. The desalted peptides were eluted in three steps with addition of 200 μ L 0.1% TFA in 50% ACN, 200 μ L of 0.1% TFA in 80% ACN, and 200 μ L of 0.1% TFA in 100% ACN, with centrifugation at 500 g/3 min between each step. The eluted peptides were vacuum-dried in a SpeedVac concentrator and stored at -20 °C.

Peptide fractionation. For the spectral library generation, 90 μ L aliquots of desalted peptide eluates corresponding to 15 μ g of starting protein material were transferred from individual samples and pooled. 800 μ g of the pooled sample was subjected to hydrophilic chromatography (HILIC) fractionation as described previously²⁸ with the use of HILIC Kinetex column (Phenomenex, USA, 2.6 μ m, 150 \times 2.1 mm, 100 Å) connected to Agilent Infinity 1260 LC system (Agilent, USA) with mobile phase (A) 50 mM ammonium formate pH 3.2 in 100% acetonitrile, and mobile phase (B) of 50 mM ammonium formate pH 3.2 in water. The fractionation was performed with isocratic 5% B for 5 min, 7 min gradient to 25% B, then 23 min gradient to 39% B followed by 5 min gradient to 55% B, isocratic 55% B for 5 min, 0.5 min gradient to 5% B and subsequently isocratic 5% B for 4.5 min. Fractions were collected every 1 min, neighboring fractions with lower signal intensities were subsequently pooled to generate a final set of 12 fractions based on peptide content (Supplementary Fig. S1). The peptides were vacuum-dried in a SpeedVac concentrator and stored at -20 °C.

LC-MS/MS analysis. All peptide mixtures were extracted into LC-MS vials by 2.5% formic acid (FA) in 50% ACN and 100% ACN with the addition of polyethylene glycol (final concentration 0.001%)³⁰ and concentrated in a SpeedVac concentrator (Thermo Fisher Scientific) and peptide concentration was estimated using small aliquot measurement using UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) online connected to Impact II system (Bruker) using the area under the total ion chromatogram curve as a total peptide quantity estimate and using external calibration curve done using HeLa (Pierce) solutions. LC-MS/MS analyses of all peptide extracts (200 ng per injection based on the estimated total peptide concentration) from 12 HILIC fractions and 16 individual samples from the original sample set used for generation of the pooled sample for HILIC fractionation were performed using nanoElute system (Bruker) directly connected to timsTOF Pro mass spectrometer (Bruker) as described previously³¹ with modifications for DDA-PASEF measurements.

MSn data were acquired in DDA-PASEF for the HILIC fractions and DIA-PASEF approach for the individual samples. m/z range of 100–1700 and 1/k0 range of $0.6\text{--}1.6 \text{ V} \times \text{s} \times \text{cm}^{-2}$ was used in both methods. DDA-PASEF method acquired 10 PASEF scans with scheduled target intensity of 20,000 and intensity threshold of 2,500. Active exclusion was set for 0.4 min with precursor reconsideration for $4 \times$ more intense precursors. DIA-PASEF method used m/z 400–1100 precursor range and equal windows size of 26 Th (incl. 1 Th overlaps) using two steps each PASEF scan and cycle time of 100 ms locked to 100% duty cycle (Supplementary File 2).

DDA data processing and spectral library generation. The DDA data from 12 HILIC peptide fractions and 16 DIA runs of individual samples used for spectral library generation were searched by Pulsar algorithm implemented in Spectronaut 16.0 software (Biognosys) against human UniProt/SwissProt database (version 2022_03 downloaded on 2022-09-23, 20,398 sequences). Enzyme specificity was set to trypsin/P, two missed cleavages were allowed, fixed modifications were set to methylthiolation (C), and variable modifications were set to oxidation (M) and acetylation (protein N-terminus). Protein, peptide and peptide-spectrum match (PSM) FDR were set to 0.01. MS1 and MS2 tolerances were set as dynamic with correction factor 1. Other parameters were set as default.

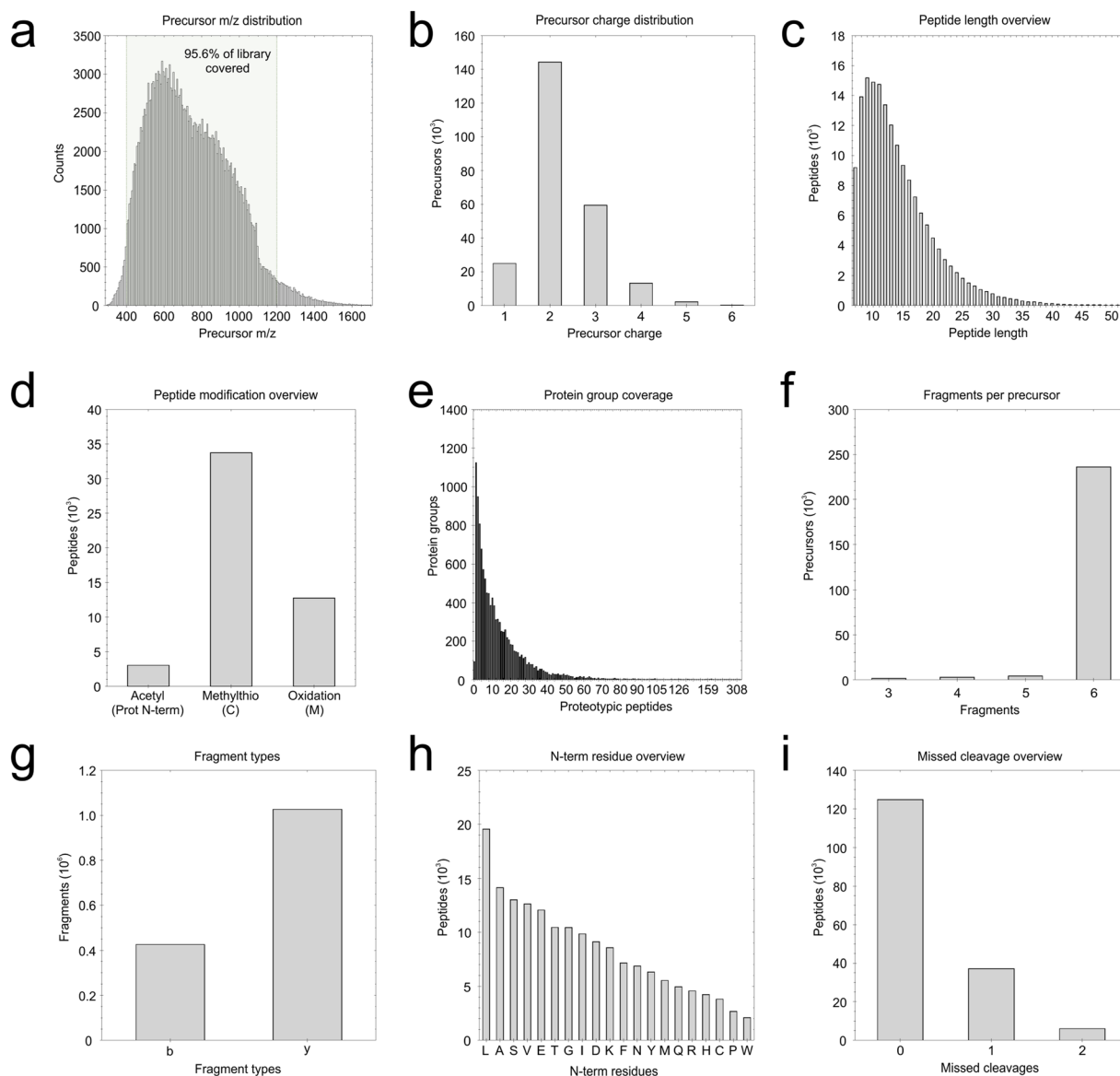


Fig. 2 Characteristics of the TNBC-specific spectral library. **(a)** m/z scale of the precursors. **(b)** Precursor charge distribution. **(c)** Peptide length distribution. **(d)** Numbers of identified modified peptides. **(e)** Distribution of numbers of proteotypic peptides that were identified per protein. **(f)** Numbers of fragments per precursor. **(g)** Fragment ion types. **(h)** N-terminal amino acid distribution. **(i)** Missed cleavage rate.

DIA data processing in Spectronaut software. The 16 TNBC samples measured in DIA-PASEF mode used for library generation were subjected to quantitative data extraction in Spectronaut 18.5 software using our hybrid library generated in Spectronaut 16.0. Precursor Qvalue cutoff and experiment protein Qvalue cutoff were set to 0.01. No missing value imputation was used. Other parameters were set as default. The results were reported from sparse profiles including peptides identified in at least one sample with Qvalue < 0.01³².

Statistical analysis. The protein profile boxplots, distribution of missing values and data completeness were visualized using DIA-Analyst tool³³. The plot of protein level coefficient of variation (CV) distribution was generated with MixProTool³⁴.

Data Records

The raw DDA-PASEF mass spectrometry proteomics data of peptide fraction measurements, Pulsar spectral library files, raw DIA-PASEF data for 16 samples used for targeted data extraction as well as Spectronaut and DIA-NN data analysis outputs have been deposited in the ProteomeXchange Consortium via the Proteomics Identifications (PRIDE) partner repository³⁵ (<http://www.ebi.ac.uk/pride/archive/>) with the dataset identifier PXD047793³⁶.

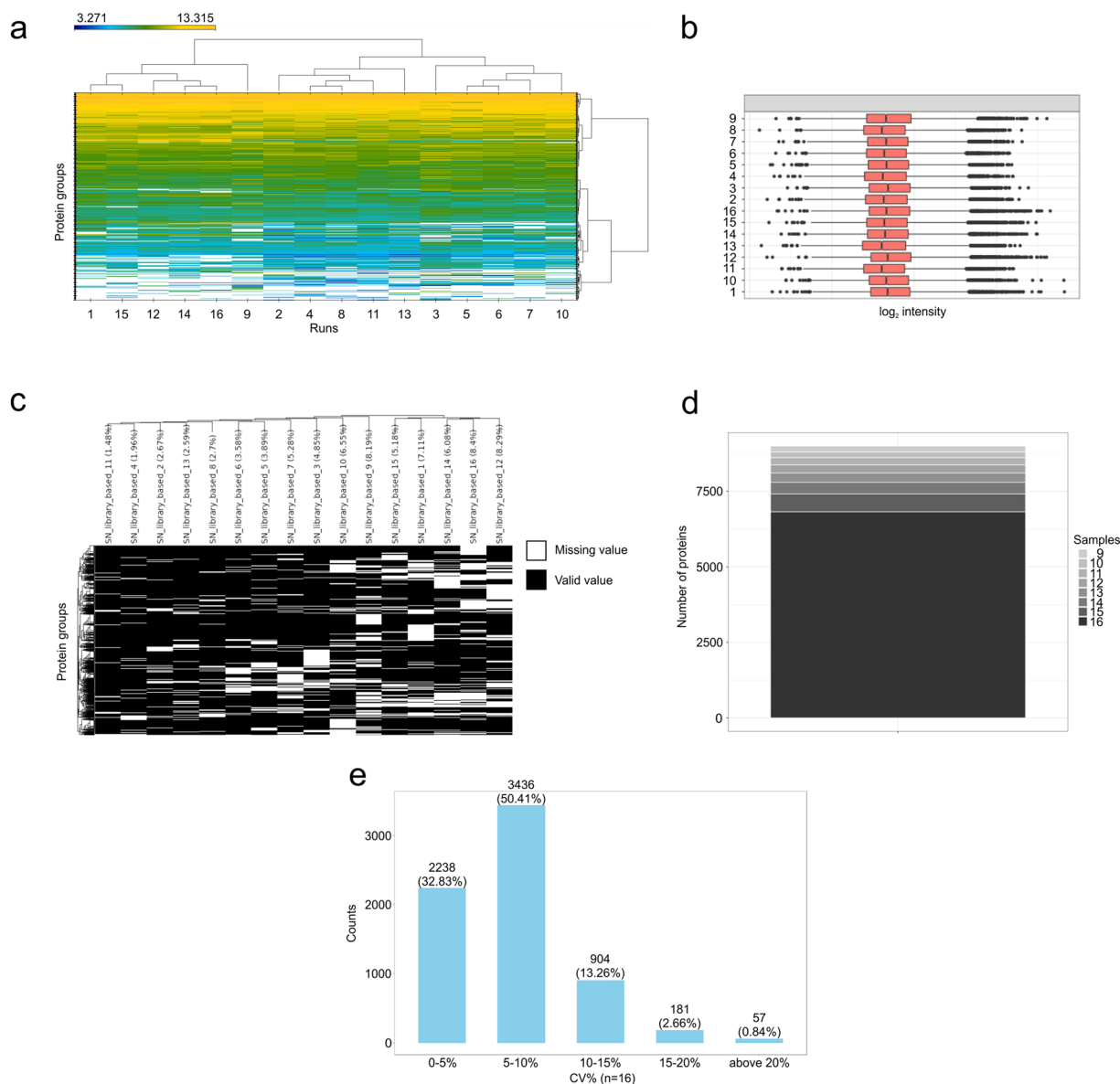


Fig. 3 The application of our TNBC library for DIA data extraction from DIA-PASEF measurements of the 16 individual TNBC samples in Spectronaut 18.5. **(a)** Heatmap displaying the sample clustering. **(b)** Protein group quantities within individual runs. **(c)** Distribution of missing values. Only proteins with at least one missing value are visualized. **(d)** Protein identification completeness. **(e)** Distribution of CVs on the protein group level.

Technical Validation

For generation of TNBC-specific library of proteomics assays, peptides from total of 105 TNBC tissues were pooled and fractionated by HILIC (Supplementary Fig. S1) and subsequently conducted to LC-MS/MS analysis on timsTOF Pro mass spectrometer in DDA-PASEF mode, whereas 16 of the individual peptide samples from the original sample group used for HILIC fractionation were measured in DIA-PASEF mode (Supplementary File 2). Spectronaut 16.0 (Biognosys) software was used to build a hybrid library of assays consisting of a Pulsar search archive (*.psar file) of 16 DIA-PASEF measurements of individual samples, and a Pulsar search archive with 12 DDA-PASEF measurements of HILIC fractions and is accessible via PRIDE together with the raw data files as well as processed data files (Supplementary File 3). This final library includes 244,464 precursors and 168,006 stripped peptide sequences corresponding to 11,564 protein groups (FDR = 0.01). For 95.6% of precursors, the range of the mass was between 400 and 1,200 m/z (Fig. 2a) with one (10.2%), two (59.0%), three (24.3%) and four (5.4%) positive charges (Fig. 2b). The peptide length varied from 7 to 52 amino acids (Fig. 2c). The modifications included oxidation on methionine for 12,777 peptides, acetylation on N-terminus for 3,055 peptides, and methylthiolation on cysteine for 33,749 peptides (Fig. 2d). Of these, minimum of two proteotypic peptides were identified for 10,347 protein groups, and at least eight proteotypic peptides were identified for 6,375 protein groups (Fig. 2e). Precursors had mostly (235,979 precursors) more than 5 fragment ions (Fig. 2f), with the most prevalent (70.7%)

y-ion series (Fig. 2g). The most frequent N-terminal amino acid was leucine (19,554 peptides) (Fig. 2h) and the protein digestion efficacy was demonstrated by 124,827 peptides with no missed cleavage (Fig. 2i).

Usage Notes

The goal of this work is the generation of the comprehensive library of assays specific for TNBC in order to support DIA proteomics research of TNBC allowing deep proteome coverage.

From the pooled and fractionated 105 TNBC samples that were used for generation of our hybrid library in Spectronaut 16.0, we selected 16 individual samples that were measured in DIA-PASEF mode and processed in the most recent software version, Spectronaut 18.5 with the use of our TNBC library. Using this approach, we identified 190,310 precursors, 140,566 stripped peptides, and 10,463 protein groups (FDR = 0.01, Supplementary file 4). The heatmap displays sample clustering, showing three potential clusters among the 16 samples (Fig. 3a). The total intensity variability between individual runs was minimal (Fig. 3b). The missing value rate varied from 1.48% to 8.4% (Fig. 3c), and 6816 protein groups were identified in all 16 samples (Fig. 3d). 83.2% and 99.1% protein groups were quantified with CV below 10% and 20%, respectively (Fig. 3e).

To compare the performance of our library with library-free approach of DIA data analysis and to compare the utilization of another widely used DIA data processing software, DIA-NN, with the Spectronaut, we show the results for the same 16 individual samples from the outcomes of library-free analysis in Spectronaut 18.5 and both library-based and library-free analyses in DIA-NN 1.8.1 in the Supplementary File 5 (the comparison results are visualized in Supplementary Figs. S2 and S3, for detailed protein group-level data please see Supplementary File 6-9).

In conclusion, in this study we present a novel proteomics assay library derived from TNBC tissues by the timsTOF technology representing the deepest coverage of TNBC proteome to date. We also report comparison of two key DIA data extraction strategies, Spectronaut and DIA-NN, and show that implementation of our library in the quantitative analysis represents a sensitive approach outperforming library-free methods. Our TNBC-specific library offers a comprehensive source for a deep proteomics characterization of TNBC tumors, leading to identification of novel therapeutic targets and precise patient stratification.

Code availability

The data analysis methods as well as software, their versions and parameters used are described in the Methods section. No custom scripts were generated in this work.

Received: 14 February 2024; Accepted: 10 July 2024;

Published online: 18 July 2024

References

1. Sporikova, Z., Koudelakova, V., Trojanec, R. & Hajduch, M. Genetic Markers in Triple-Negative Breast Cancer. *Clin Breast Cancer* **18**, e841–e850 (2018).
2. Won, K.-A. & Spruck, C. Triple-negative breast cancer therapy: Current and future perspectives (Review). *Int J Oncol* **57**, 1245–1261 (2020).
3. Meng, X., Cai, Y., Chang, X. & Guo, Y. A novel conditional survival nomogram for monitoring real-time prognosis of non-metastatic triple-negative breast cancer. *Front Endocrinol (Lausanne)* **14**, 1119105 (2023).
4. Gluz, O. *et al.* Triple-negative breast cancer—current status and future directions. *Ann Oncol* **20**, 1913–1927 (2009).
5. Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E. & Gianni, L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol* **13**, 674–690 (2016).
6. Maqbool, M., Bekele, F. & Fekadu, G. Treatment Strategies Against Triple-Negative Breast Cancer: An Updated Review. *Breast Cancer (Dove Med Press)* **14**, 15–24 (2022).
7. Kreike, B. *et al.* Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res* **9**, R65 (2007).
8. Kelstrup, C. D. *et al.* Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J Proteome Res* **17**, 727–738 (2018).
9. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11**, O111.016717 (2012).
10. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
11. Faktor, J., Sucha, R., Paralova, V., Liu, Y. & Bouchal, P. Comparison of targeted proteomics approaches for detecting and quantifying proteins derived from human cancer tissues. *Proteomics* **17**, (2017).
12. Barkovits, K. *et al.* Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. *Mol Cell Proteomics* **19**, 181–197 (2020).
13. Shao, W. & Lam, H. Tandem mass spectral libraries of peptides and their roles in proteomics research. *Mass Spectrom Rev* **36**, 634–648 (2017).
14. Tsou, C.-C. *et al.* DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* **12**, 258–264 (2015). 7 p following 264.
15. Yang, Y. *et al.* In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics. *Nat Commun* **11**, 146 (2020).
16. Muntel, J. *et al.* Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. *J Proteome Res* **14**, 4752–4762 (2015).
17. Lapcik, P. *et al.* A large-scale assay library for targeted protein quantification in renal cell carcinoma tissues. *Proteomics* e2100228 <https://doi.org/10.1002/pmic.202100228> (2021).
18. García-Adrián, S. *et al.* Molecular characterization of triple negative breast cancer formaldehyde-fixed paraffin-embedded samples by data-independent acquisition proteomics. *Proteomics* **22**, e2100110 (2022).
19. Meng, X., Liu, D. & Guan, Y. Advances in the application of label-free quantitative proteomics techniques in malignancy research. *Biomed Chromatogr* **37**, e5667 (2023).
20. Meier, F. *et al.* Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics* **17**, 2534–2545 (2018).
21. Meier, F. *et al.* diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat Methods* **17**, 1229–1236 (2020).

22. Zhang, F., Ge, W., Ruan, G., Cai, X. & Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **20**, e1900276 (2020).
23. Bruderer, R. *et al.* Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* **14**, 1400–1410 (2015).
24. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* **17**, 41–44 (2020).
25. Skowronek, P. & Meier, F. High-Throughput Mass Spectrometry-Based Proteomics with dia-PASEF. *Methods Mol Biol* **2456**, 15–27 (2022).
26. Demichev, V. *et al.* dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nat Commun* **13**, 3944 (2022).
27. Muntel, J. *et al.* Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol Omics* **15**, 348–360 (2019).
28. Bouchal, P. *et al.* Breast Cancer Classification Based on Proteotypes Obtained by SWATH Mass Spectrometry. *Cell Rep* **28**, 832–843. e7 (2019).
29. Bouchal, P. *et al.* Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. *J Proteome Res* **8**, 362–373 (2009).
30. Stejskal, K., Potěšil, D. & Zdráhal, Z. Suppression of peptide sample losses in autosampler vials. *J Proteome Res* **12**, 3057–3062 (2013).
31. Řihová, K. *et al.* Caspase-9 Is a Positive Regulator of Osteoblastic Cell Migration Identified by diaPASEF Proteomics. *J Proteome Res* <https://doi.org/10.1021/acs.jproteome.3c00641> (2024).
32. Lou, R. *et al.* Benchmarking commonly used software suites and analysis workflows for DIA proteomics and phosphoproteomics. *Nat Commun* **14**, 94 (2023).
33. Monash Proteomics and Metabolomics Platform & Monash Bioinformatics Platform, Monash University. DIA-Analyst: An interactive web-platform to analyze and visualize proteomics data preprocessed with Spectronaut and DIA-NN. <https://analyst-suites.org/apps/dia-analyst/>.
34. Wang, S., Zheng, W., Hu, L., Gong, M. & Yang, H. MixProTool: A Powerful and Comprehensive Web Tool for Analyzing and Visualizing Multigroup Proteomics Data. *J Comput Biol* **25**, 1123–1127 (2018).
35. Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**, D543–D552 (2022).
36. Lapcik, P. *et al.* A hybrid DDA/DIA-PASEF-based assay library for a deep proteotype characterization of triple-negative breast cancer. *PRIDE Archive* <https://identifiers.org/pride.project:PXD047793> (2023).

Acknowledgements

We thank Sarka Hrachovinova (Masaryk University, Faculty of Science) for drafting the general introduction on breast cancer. This work was primarily supported by Ministry of Health of the Czech Republic (project NU22-08-00230), all rights reserved. CIISB, Instruct-CZ Centre of Instruct-ERIC EU consortium, funded by MEYS CR infrastructure project LM2023042 and European Regional Development Fund-Project “UP CIISB” (No. CZ.02.1.01/0.0/0.0/18_046/0015974), is gratefully acknowledged for the financial support of the measurements at the CEITEC Proteomics Core Facility (by DP). Supported by the project BBMRI.CZ no. LM2023033. Supported by the project National Institute for Cancer Research (Programme EXCELES, ID Project No. LX22NPO5102)—Funded by the European Union—Next Generation EU.

Author contributions

P. Bouchal designed the project. R.N. collected samples. P.L., J.L. and P. Bouchalova processed the samples. D.P. performed LC-MS/MS measurements. P. Bouchal, K.S. and P.L. conducted proteomics data analysis. P.L. drafted the manuscript. P. Bouchal supervised the project, acquired funding and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03632-2>.

Correspondence and requests for materials should be addressed to Pavel Bouchal.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024