MASS SPECTROMETRY-BASED ANALYSIS OF HUMAN TISSUE SAMPLES IN PROSTATE AND LUNG CANCER RESEARCH

PhD thesis

Simon Nándor Sugár

Doctoral School of Pharmaceutical Sciences
Semmelweis University

Supervisor: Lilla Turiák, PharmD, PhD
Official reviewers: Éva Csősz, PhD
Borbála Dalmadiné Kiss, PhD

Head of the Complex Examination Committee:
Zoltán Benyó, MD, DSc

Members of the Complex Examination Committee:
Krisztina Ludányi, PhD
Gitta Vácziné Schlosser, PhD

Budapest
2023

1. Introduction

Prostate cancer (PCa) research—including proteomics and *N*-glycosylation—is currently focused on two main topics: finding novel diagnostic markers to replace the prostate specific antigen (PSA) blood test, and prognostic markers that accurately predict aggressive PCa.

Replacing, or improving the accuracy of the PSA blood test has been long on the forefront of PCa proteomics research, as it has low specificity towards PCa causing overdiagnosis and overtreatment. For this reason, there is significant effort directed towards finding novel biomarkers. There are several promising putative markers—including several different protein panels—currently under investigation, however, none of them have been implemented into clinical practice in a widespread manner yet. Although not as central a research field as the two mentioned above, improving therapeutic options for advanced PCa (the treatment of low-risk PCa is relatively well-established) along with drug resistance and therapeutic response markers is an important and significant part of PCa research as well.

Currently, lung cancer (LC) proteomics studies are mostly focused on discovering novel diagnostic, prognostic, and predictive molecular markers; and potential therapeutic targets to improve currently available options. The analyzed sample types include cell

lines, body fluids (blood, urine, saliva etc.,), and fresh frozen or formalin-fixed paraffin embedded (FFPE) tissue.

The proteomic analysis of tissue samples is central to understanding LC biology, as it enables the selective study of certain groups and subtypes of the disease in an in vivo setting, and also intratumoral heterogeneity. Furthermore, there are multiple tissue-based markers currently in clinical use, several of them proteins: immunohistochemical markers (e.g., thyroid transcription factor 1, napsin A for adenocarcinoma; and p40, and cytokeratin 5/6 for squamous cell carcinoma), genomic markers (e.g., epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), or Kirsten rat sarcoma viral oncogene homolog (KRAS)), and immunotherapy markers (e.g., cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4), and programmed death-ligand 1 receptor (PD-1)).

2. Objectives

In the thesis two separate research projects are summarized: the proteomic and *N*-glycoproteomic analysis of cancerous and healthy prostate tissue microarray (TMA) samples; and the proteomic analysis of different types of lung cancer tissue. Although cancer is at the center of both, their focus is very different—which is in line with current cancer-specific clinical demands and challenges.

The first project—the characterization of prostate TMA samples—was part of the "Mass spectrometry-based identification of prostate cancer biomarkers from tissue microarrays" project funded by the National Research, Development and Innovation Office (OTKA PD 121187). The aim was the integration of previous research and method development, and its application to the analysis of prostate TMAs to identify molecular alterations with potential diagnostic value. The most important prerequisites were the successful application of on-surface tryptic digestion to prostate TMAs that contain very limited sample amounts in a proteomic pilot study comparing healthy and cancerous samples, and the optimization and application of a glycopeptide enrichment method—acetone precipitation—to similar sample amounts. The key objectives of the project were: successful combination of on-surface digestion and acetone precipitation to obtain proteomic and *N*-glycoproteomic information in parallel from the same TMA

sample, and the investigation of different grades of PCa for a better understanding of disease progression.

The second project—the proteomic analysis of lung tissue sections—was part of the "Identification of predictive biomarkers for the prognosis of targeted biological therapy in lung cancer" project funded by the National Research, Development and Innovation Office (OTKA FK 131603). Here, the objective was to identify dysregulated biological processes by analyzing and comparing the four main types of LC (small cell lung cancer (SCLC), adenocarcinoma (AC), squamous cell carcinoma (SqCC), and large cell carcinoma (LCC)) with their adjacent normal regions and each other by proteomics using on-surface tryptic digestion of FFPE tissue, enabling a more precise diagnosis and classification of lung cancer.

3. Methods

For the first project, four different TMA slides were purchased from US Biomax: BNS19011, PR481, PR483c, PR633. All of them contained FFPE cores with a diameter of 1.5 mm and a thickness of 5 μm. The specification sheets are available at US Biomax's website with information about each core including age, pathological Grade, Stage, and Gleason Score. Each TMA core contains on average approximately 1 μg protein.

For the second project, formalin-fixed paraffin-embedded AC, SqCC, LCC, SCLC (n = 10, 9, 10, 9 respectively) and tumor-adjacent normal tissue sections (n = 8, 8, 8, 9 respectively) were analyzed. The work was approved by the Medical Research Council (TUKEB permit number: IV/2567-4/2020/EKU). Inclusion criteria for patients were the following: fresh LC cases with resection specimens, while also keeping in mind that histological groups should have similar sizes (the aim was 10 patients in each group). FFPE tissue sections with a thickness of 10 μm were obtained from the departmental archive of the Department of Pathology, University of Pécs, Hungary.

The main steps of the analysis were as follows: on-surface proteolytic digestion, reversed-phase purification, glycopeptide enrichment by acetone precipitation (only for prostate TMAs), nanoUHPLC-MS/MS analysis, and data processing and analysis.

The on-surface digestion consisted of several steps, the baking, deparaffinization, and antigen retrieval of tissue samples followed by reduction and alkylation, then proteolysis using LysC-trypsin and trypsin enzymes. For reversed-phase purification C18 solid phase extraction (SPE) columns were used. This was followed by glycopeptide enrichment for prostate TMA samples, after which the pellet and supernatant fractions (containing most glycopeptides and peptides, respectively) were analyzed separately. For the nanoUHPLC-MS/MS analysis a Maxis II QTOF instrument equipped with a CaptiveSpray nanoBooster ion source coupled to a Dionex UltiMate 3000 RSLCnano system was used. For protein and *N*-glycopeptide identification Byonic, for quantitation MaxQuant and GlycoPattern software was used. Data curation, analysis, statistical evaluation, and visualizations were done in R using RStudio. For protein-protein interaction (PPI) networks STRING and gene set enrichment analysis (GSEA) were used.

4. Results

4.1. Prostate Cancer Proteomics and *N*-glycosylation

During the project aiming at the proteomic and *N*-glycoproteomic characterization of cancerous and normal prostate tissue, 95 TMA biopsy samples were analyzed. Among these, there were 9 grade 1 (G1), 16 grade 2 (G2), 24 grade 3 (G3), and 46 normal.

To investigate differences between normal and cancerous tissue, t-tests were used with a 0.05 FDR. In the proteomics dataset 123 proteins were found to be differentially expressed, this included 72 proteins overexpressed and 51 proteins underexpressed in PCa. Among these, 14 showed a fold-change over 2, while 27 displayed a fold-change under 0.5. Following the identification of proteins with statistically significant changes, functional enrichment analysis was performed in STRING; most of the underexpressed proteins were associated with cellular component organization, while the overexpressed proteins were predominantly affiliated with metabolic processes. In the glycoproteomics dataset, 7 glycopeptides were found with significantly different abundances between the normal and PCa groups each carrying biantennary, fucosylated complex-type glycans with different levels of galactosylation and sialylation. Differences were also detected between normal and PCa tissues when comparing the levels of sialylation, fucosylation, and galactosylation at distinct glycosites.

To uncover molecular alterations among pathological grades and normal tissue, ANOVA was used with FDR controlled at 0.05. In the proteomics dataset, 75 proteins were identified with significant changes among the various PCa grades and normal tissue. Hierarchical clustering based on Spearman's correlation revealed two distinct groups among these proteins: 40 proteins were mostly upregulated, while 35 downregulated with cancer progression. In the glycoproteomics dataset, 4 glycopeptides were identified with significantly different abundances among different grades and healthy tissue. Furthermore, regarding glycosites, the degree of fucosylation on collagen type VI alpha 2 chain (CO6A2) N785 was found to be different between the three Grade groups and Normal tissue.

4.2. Lung Cancer Proteomics

To characterize the proteomic differences between the different types of lung cancer (AC, SqCC, LCC, SCLC), 71 samples were analyzed. These were derived from FFPE tissue sections taken from individuals suffering from either of the four different types of LC. From the tissue sections, samples were taken from both the cancerous and the cancer-adjacent normal regions, 8–10 samples belong to each sample group.

Two-group comparisons were performed separately for all 4 LC types. This revealed that there are 78 proteins differentially expressed in all four LC types compared to adjacent tissue. The

majority of alterations occurred in a group-specific manner: 61 proteins were differentially expressed only in AC, while 201, 119, and 44 proteins only in SCLC, SqCC, and LCC, respectively.

Proteins differentially expressed between the cancerous tissue regions of the four types of LC were identified using multiple sample and two-sample comparisons, resulting in 571 proteins with altered expression. None of the proteins showed altered expression in all pairwise comparisons, but 23 of them showed differences in 5 out of 6 comparisons. Furthermore, there were several proteins with changes in expression in only one LC type, separating it from all the others. Hierarchical clustering based on the 571 differentially expressed proteins revealed that the molecular profiles of the different types of LC were markedly different, and the clustering of samples—except for 2—was in agreement with the pathological classification.

To identify the dysregulated biological processes in the tumorous regions compared to adjacent tissue, pre-ranked GSEA was applied. This was performed separately for all four LC types (based on effect sizes), then the results were compared. The enriched gene sets could be separated into several major groups of biological processes: extracellular matrix organization, assembly, regulation and adhesion; signaling cascades (e.g., Ca2+ dependent signaling, RHO signaling); processes involved in protein synthesis (transcription, translation, DNA, and RNA related processes);

humoral and cell-mediated immune system processes; and transport processes (e.g., vesicular transport, endocytosis). The processes dysregulated in LC were similar in all types; however, there were several type-specific differences, especially between SCLC and non-small cell lung cancer (NSCLC).

5. Conclusions

The results of the proteomic and *N*-glycoproteomic analysis of prostate TMAs indicate that alterations between PCa and Normal tissue glycosylation occur primarily at the glycosite level, while overall glycosylation may remain unaffected. Furthermore, altered glycosylation does not necessarily indicate a change in protein expression. The glycoproteins with altered *N*-glycosylation were all secreted either into the bloodstream or the extracellular matrix, and most of them are characterized as an unfavorable prognostic cancer marker by the Pathology Atlas. As altered protein glycosylation in cancer has been proven to be nonrandom, this suggests that further investigation—especially regarding cancer specificity—of these potential prognostic markers and identification of their exact roles is reasonable and could lead to further advancement in understanding the function of *N*-glycosylation in cancer development, and PCa prognosis.

The proteomic analysis of FFPE tissue samples from four different types of lung cancer resulted in the identification of several disrupted biological processes in all investigated cancer types, such as the degradation of the basement membrane and suppression of the complement and coagulation cascade, as well as the activation of the MTORC1 signaling pathway. The differences between SCLC and NSCLC were larger than between the three distinct NSCLC subtypes. Dysregulated pathways

differentiating SCLC from NSCLCs include suppressed regulation of cell adhesion, actin filament-based processes, and calcium ion binding. Overexpression of splicing factors and heterogeneous nuclear ribonucleoproteins suggests that biological processes related to splicing are more affected in SCLC. Additionally, several proteins showed differential expression only in one LC type, such as the overexpression of fascin actin-bundling protein 1 in SqCC or the downregulation of nidogen-2 in SCLC. These results correlated well with previous studies analyzing individual NSCLC types and tumor adjacent tissue. The specific molecular signatures connected to LC in general and to specific LC types described in detail might be attractive targets for further in-depth investigations and may have potential diagnostic and prognostic value.

6. Bibliography of the candidate's publications

Publications related to the dissertation:

1: Sugár, S., Tóth, G., Bugyi, F., Vékey, K., Karászi, K., Drahos, L., & Turiák, L. (2021). Alterations in protein expression and site-specific N-glycosylation of prostate cancer tissues. *Scientific reports*, 11(1), 1-12, 15886.

IF (2021): 4.997

2: Sugár, S., Bugyi, F., Tóth, G., Pápay, J., Kovalszky, I., Tornóczky, T., Drahos, L., & Turiák, L. (2022). Proteomic Analysis of Lung Cancer Types—A Pilot Study. *Cancers*, *14*(11), 2629.

IF (2021): 6.575

Publications unrelated to the dissertation:

1: Turiák, L., Sugár, S., Ács, A., Tóth, G., Gömöry, Á., Telekes, A., Vékey, K., & Drahos, L. (2019). Site-specific N-glycosylation of HeLa cell glycoproteins. *Scientific reports*, *9*(1), 1-11.

2: Sugár, S., Turiák, L., Vékey, K., & Drahos, L. (2020). Widespread presence of bovine proteins in human cell lines. *Journal of Mass Spectrometry*, *55*(7), e4464.

3: Tóth, G., Vékey, K., Sugár, S., Kovalszky, I., Drahos, L., & Turiák, L. (2020). Salt gradient chromatographic separation of chondroitin sulfate disaccharides. *Journal of Chromatography A*, *1619*, 460979.

4: Bokka, R., Ramos, A. P., Fiume, I., Manno, M., Raccosta, S., Turiák, L., Sugár, S., Adamo, G., Csizmadia, T., & Pocsfalvi, G. (2020). Biomanufacturing of Tomato-Derived Nanovesicles. *Foods*, *9*(12), 1852.

5: Tóth, G., Bugyi, F., Sugár, S., Mitulović, G., Vékey, K., Turiák, L., & Drahos, L. (2020). Selective TiO2 Phosphopeptide Enrichment of Complex Samples in the Nanogram Range. *Separations*, *7*(4), 74.

6: Tóth, G., Pál, D., Sugár, S., Kovalszky, I., Dezső, K., Schlosser, G., Drahos, L., & Turiák, L. (2022). Expression of glycosaminoglycans in cirrhotic liver and hepatocellular carcinoma—a pilot study including etiology. *Analytical and bioanalytical chemistry*, *414*(13), 3837-3846.

7: Sallai, I., Marton, N., Szatmári, A., Kittel, Á., Nagy, G., Buzás, E. I., Khamari, D., Komlósi, Zs., Kristóf, K., Drahos, L., Turiák, L., Sugár, S., Veres, D. S., Kendoff, D., Zahár, Á., & Skaliczki, G. (2022). Activated polymorphonuclear derived extracellular vesicles are potential biomarkers of periprosthetic joint infection. *Plos one*, *17*(5), e0268076.

8: Luparello, C., Branni, R., Abruscato, G., Lazzara, V., Sugár, S., Arizza, V., Mauro, M., Di Stefano, V., & Vazzana, M. (2022). Biological and proteomic characterization of the anti-cancer potency of aqueous extracts from cell-free coelomic fluid of Arbacia lixula sea urchin in an in vitro model of human hepatocellular carcinoma. *Journal of Marine Science and Engineering*, *10*(9), 1292.

9: Tóth, G., Sugár, S., Pál, D., Fügedi, K. D., Drahos, L., Schlosser, G., Oláh, C., Reis, H., Kovalszky, I., Szarvas, T., & Turiák, L. (2022). Glycosaminoglycan Analysis of FFPE Tissues from Prostate Cancer and Benign Prostate Hyperplasia Patients Reveals Altered Regulatory Functions and Independent Markers for Survival. *Cancers*, *14*(19), 4867.

10: Hevér, H., Nagy, K., Xue, A., Sugár, S., Komka, K., Vékey, K., Drahos, L., & Révész, Á. (2022). Diversity Matters: Optimal Collision Energies for Tandem Mass Spectrometric Analysis of a Large Set of N-Glycopeptides. *Journal of Proteome Research*, *21*(11), 2743-2753.

11: Tóth, G., Sugár, S., Balbisi, M., Molnár, B. A., Bugyi, F., Fügedi, K. D., Drahos, L., & Turiák, L. (2022). Optimized Sample Preparation and Microscale Separation Methods for High-Sensitivity Analysis of Hydrophilic Peptides. *Molecules*, *27*(19), 6645.

12: Sugár, S., Drahos, L., & Vekey, K. (2023). Quantitative proteomics I.: Concept, design, and planning of quantitative proteomics experiments. Journal of Mass Spectrometry, 58(4), e4907.

1**3:** Pál, D., Tóth, G., Sugár, S., Fügedi, K. D., Szabó, D., Kovalszky, I., Papp, D., Schlosser, G., Tóth, C., Tornóczky, T., Drahos, L., & Turiák, L. (2023). Compositional analysis of glycosaminoglycans in different lung cancer types—a pilot study. International Journal of Molecular Sciences, 24(8), 7050.

14: Balbisi, M., Sugár, S., Schlosser, G., Szeitz, B., Fillinger, J., Moldvay, J., Drahos, L., Szász, A. M., Tóth, G., & Turiák, L. (2023). Inter-and intratumoral proteomics and glycosaminoglycan characterization of ALK rearranged lung adenocarcinoma tissues: a pilot study. Scientific Reports, 13(1), 6268.

15: Abruscato, G., Chiarelli, R., Lazzara, V., Punginelli, D., Sugár, S., Mauro, M., Librizzi, M., Di Stefano, V., Arizza V., Vizzini, A., Vazzana, M., & Luparello, C. (2023). In Vitro Cytotoxic Effect of Aqueous Extracts from Leaves and Rhizomes of the Seagrass Posidonia oceanica (L.) Delile on HepG2 Liver Cancer Cells: Focus on Autophagy and Apoptosis. Biology, 12(4), 616.