INVESTIGATING MOLECULAR HETEROGENEITY IN LUNG CANCER USING PROTEOMIC AND BIOINFORMATIC APPROACHES

PhD thesis

Beáta Szeitz

Doctoral College of Semmelweis University

Pathology and Oncology Sciences Division

Supervisor: Attila Marcell Szász, MD PhD habil
Consultants: Melinda Rezeli, PhD
Péter Horvatovich, PhD
Official reviewers: Sándor Spisák, PhD
Katalin Dezső, PhD habil

Head of the Complex Examination Committee:
Miklós Tóth, MD DSc
Members of the Complex Examination Committee:
Lilla Turiák, PhD
Tibor Glasz, MD PhD

Budapest
2024

1. INTRODUCTION

Lung cancer is a heterogeneous disease both in terms of its clinicopathological and molecular features, making its diagnosis and treatment complex and challenging in the clinic. Proteomics, the large-scale investigation of proteins, provides valuable insights into tumor biology, and aids the identification of diagnostic, predictive, prognostic, and therapeutic markers. My doctoral thesis focuses on the molecular heterogeneity of two less frequent lung cancer types, small cell lung cancers (SCLCs) and anaplastic lymphoma kinase (*ALK*)-rearranged lung adenocarcinomas (LADCs), investigated using mass spectrometry (MS)-based proteomics and complemented with results from transcriptome-based findings.

SCLC accounts for only 15% of all lung cancer cases. So far, it has been treated as a single disease in the clinic. However, recent molecular profiling studies have revealed significant

heterogeneity within SCLC, leading to the proposal of new molecular subtypes based on neuroendocrine (NE) features and the expression patterns of transcription factors *ASCL1*, *NEUROD1* (SCLC-A and N, NE types) *POU2F3* and *YAP1* (SCLC-P and Y, non-NE types). A protein-level understanding of these subtypes is necessary to develop more effective, targeted therapies for SCLC.

LADC is the most prevalent histological type of non-small cell lung cancers. Approximately 3-5% of LADC cases in the Caucasian population are driven by *ALK* gene rearrangements, which can be successfully targeted with ALK inhibitors. Despite the initial success, however, patients often relapse due to cancer cells acquiring resistance to treatment. Therapeutic resistance is strongly associated with intratumoral heterogeneity, which needs further investigation with molecular studies.

2. OBJECTIVES

1. Proteomic study of SCLC (*study I*): To perform a proteomic analysis of cell lines derived from human SCLCs, including the analysis of both the cell pellet (CP) and culture media (CM). To investigate the proteomic differences between cell lines from SCLC-A/N/P/Y subtypes, integrated with results from existing transcriptomic datasets. To list potential diagnostic or therapeutic markers for the subtypes as well as insights into the subtypes' specific pathway-level features that may influence therapy response.
2. Multi-omic study of *ALK*-rearranged LADC (*study II*): To perform a spatial multi-omic characterization of treatment-naïve *ALK*-rearranged LADCs, by utilizing both bottom-up proteomics and NanoString GeoMx gene expression profiling. To describe the molecular characteristics of tumor regions with distinct

histopathological features, including differences in morphology, immune infiltration, stroma and mucin content, as well as to provide the main contributors to molecular heterogeneity within tumors which may influence patient outcome and therapy response.

3. METHODS

A total of 26 human SCLC cell lines were cultured for proteomic analysis. Subtype was determined via quantitative polymerase chain reaction (qPCR). Both the pellets and media were processed and subjected to label-free bottom-up proteomic analysis (digestion with a Lys-C/Trypsin mix, Dionex UltiMate 3000 RSLCnano UPLC system, Q Exactive HF-X MS, Proteome Discoverer v2.4).

The formalin-fixed paraffin-embedded, primary LADC samples were obtained from the National Korányi Institute of Pulmonology in Hungary. *ALK* positivity was determined at the 2nd

Department of Pathology, Semmelweis University, Hungary using immunohistochemistry and fluorescence in situ hybridization. Ethical approval was granted by the Medical Research Council of Hungary (2521-0/2010-1018EKU, 510/2013, 52614-4-213EKU). Histopathological evaluation was conducted after hematoxylin and eosin staining on the selected proteomic and transcriptomic regions of interest (pROIs and tROIs, respectively). The pROIs (2-6 regions per case) were subjected to label-free bottom-up proteomic analysis (digestion with a Trypsin/Lys-C mix, Dionex Ultimate 3000 RSLC nanoUHPLC, Bruker Maxis II Q-TOF, Byonic v4.2.10 and MaxQuant v1.6.17). The tROIs (12 regions per case) were analyzed with the NanoString GeoMx Digital Spatial Profiler (Illumina NovaSeq 6000, Nanostring DnD pipeline and GeoMx software v2.2.0.122).

Further post-processing and analysis steps were mainly performed with R v4.2.0. The main

statistical and bioinformatic steps were consensus clustering, single-sample score calculations, differential expression analyses, and pathway overrepresentation or gene set enrichment analyses. Nominal or Benjamini-Hochberg adjusted $p < 0.05$ was considered statistically significant.

4. RESULTS

4.1 SCLC study

Label-free proteomic analysis resulted in 8405 and 5408 proteins quantified across minimum 80% of the SCLC cell lines in the CP and CM data respectively. The subtype-defining transcription factors (*ASCL1*, *NEUROD1*, *POU2F3* and *YAP1*) showed increased protein-level expression in the cell lines classified into their corresponding subtypes (SCLC-A/N/P/Y, n = 8, 7, 4 and 7, respectively). Unsupervised consensus clustering of the CP data categorized the cell lines into four subgroups, which groups were in alignment with the cell lines'

predefined qPCR-based subtypes. In contrast to CP data, the CM data was more strongly affected by the culture type. The subtypes' NE characteristics were confirmed by the relative protein expression of NE and non-NE markers.

A total of 367 and 34 proteins showed subtype-specific expression profile in the CP and CM data, respectively. In sum, 33, 57, 32 and 271 proteins were assigned to subtypes SCLC-A, -N, -P and Y, respectively. Notably, 22 of these proteins were identified in the CM data and are also detectable in human blood plasma, and 6 proteins are druggable.

Unique pathway-level patterns were also identified for the subtypes. Most notably, SCLC-A was characterized by the overexpression of oxidative phosphorylation (OXPHOS) and respiratory chain elements. SCLC-N showed increased activity in DNA replication and transcription. Upregulation of neurotrophin

signaling pathway members was characteristic of SCLC-P cell lines. Lastly, SCLC-Y formed the most distinct subgroup, with unique upregulation of pathways such as cytokine-mediated signaling and inflammatory response, MAPK cascade or extracellular matrix (ECM) organization.

4.2 *ALK*-rearranged LADC study

The spatial profiling of the seven ALK-rearranged LADCs resulted in a total of 1811 genes and 1154 proteins being quantified across minimum 80% of the tROIs and pROIs respectively. A modest positive correlation between the two molecular layers was detected (median Pearson's r = 0.43 and 0.24 at gene and pathway levels, respectively). Unsupervised consensus clustering of both the pROIs and tROIs resulted in clusters mainly forming according to patients.

Tumors and normal adjacent tissues (NATs) showed clearly distinct expression profiles, both at pROI and tROI level. Of note, we observed the

upregulation of GPX1 in ALK-rearranged LADCs compared to NATs, which trend was not detectable in *ALK*-negative LADCs. Pathway-level differences between tumors and NATs showed known cancer hallmarks, such as impaired glycolysis, translation, or receptor tyrosine kinase signaling. Gene/protein expression and pathway-level differences could also be assigned to differences in immune infiltration patterns and mucin/stroma scores across the ROIs. Importantly, upregulation of immune system-associated pathways was observable in the tROI data with increasing immune score, and both higher mucin and stroma score resulted in upregulation of ECM organization and epithelial-mesenchymal transition (EMT)-related proteins.

At both pROI and tROI level, members of the ECM-associated pathways and EMT (and most notably, FN1) exhibited substantial gene and protein expression variability within tumors.

5. CONCLUSIONS

Study I:

1. The recently defined, qPCR-based SCLC-A/N/P/Y classification system in SCLC was confirmed at the protein level by analyzing 26 human SCLC cell lines via MS-based proteomics.
2. Proteomics uncovered a list of potential proteins that display subtype-specific expression patterns, including several potential IHC- and blood-based markers and therapeutic targets.
3. Furthermore, the subtypes also demonstrated unique signatures at the pathway-level, such as upregulated OXPHOS in SCLC-A, DNA replication in SCLC-N, neurotrophin signaling in SCLC-P and EMT in SCLC-Y.
4. The *YAP1*-driven subtype was clearly distinguishable at the protein level, and proteomics provided an abundance of potential

markers of this subtype compared to the other subtypes.

Study II:

5. The spatial proteomic and transcriptomic profiling of treatment-naïve *ALK*-rearranged LADCs was performed, where the two molecular layers exhibited modest correlation with each other.
6. Both proteomics and transcriptomics indicated that molecular heterogeneity across tumors was more prominent than molecular heterogeneity within tumors.
7. The spatial multi-omic analysis uncovered potential biomarkers and dysregulated pathways linked to tumors or NATs, varying levels of immune infiltration, mucin, and stroma content. Notably, overexpression of GPX1 in tumors compared to NATs may be detectable only in the *ALK*-rearranged subset of LADCs but not in non-*ALK*-driven LADCs.

8. Within tumors, heterogeneity in terms of the expression of ECM elements, such as FN1, was observed at proteome and transcriptome level.

6. BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

6.1 Publications related to the subjects of the thesis

Szeitz B, …, Szász AM. Spatially Resolved Proteomic and Transcriptomic Profiling of Anaplastic Lymphoma Kinase-Rearranged Pulmonary Adenocarcinomas Reveals Key Players in Inter- and Intratumoral Heterogeneity. Int J Mol Sci. 2023;24(14):11369. **IF: 5.6** (expected)

Szeitz B*, Megyesfalvi Z*, …, Döme B, Schelch K, Rezeli M. In-depth proteomic analysis reveals unique subtype-specific signatures in human small-cell lung cancer. Clin Transl Med. 2022;12(9):e1060. **IF: 10.6**

**Shared first authors.*

6.2 Other publications

Solta A, …, **Szeitz B**, …, Dome B, Schelch K. Entinostat Enhances the Efficacy of Chemotherapy in Small Cell Lung Cancer Through S-phase Arrest and Decreased Base Excision Repair. Clin Cancer Res. 2023;29(22):4644-59. **IF: 11.5** (expected)

Tisza A, …, **Szeitz B**, …, Megyesfalvi Z, Dome B. Laser ablation-inductively coupled plasma-mass spectrometry analysis reveals differences in chemotherapeutic drug distribution in surgically resected pleural mesothelioma. Br J Clin Pharmacol. 2023;89(11):3364-74. **IF: 3.4** (expected)

Balbisi M, …, **Szeitz B**, …, Turiák L. Inter- and intratumoral proteomics and glycosaminoglycan characterization of ALK rearranged lung adenocarcinoma tissues: a pilot study. Sci Rep. 2023;13(1):6268. **IF: 4.6** (expected)

Woldmar N, …, **Szeitz B**, …, Rezeli M, Döme B. Proteomic analysis of brain metastatic lung adenocarcinoma reveals intertumoral heterogeneity

and specific alterations associated with the timing of brain metastases. ESMO Open. 2023;8(1):100741. **IF: 7.3** (expected)

Megyesfalvi Z, …, **Szeitz B**, …, Laszlo V, Dome B. Expression patterns and prognostic relevance of subtype-specific transcription factors in surgically resected small-cell lung cancer: an international multicenter study. J Pathol. 2022;257(5):674-86. **IF: 7.3**

Szeitz B, …, Tőkés AM. Investigating the Prognostic Relevance of Tumor Immune Microenvironment and Immune Gene Assembly in Breast Carcinoma Subtypes. Cancers (Basel). 2022;14(8). **IF: 5.2**

Szadai L, Velasquez E, **Szeitz B**, …, Németh IB, Marko-Varga G. Deep Proteomic Analysis on Biobanked Paraffine-Archived Melanoma with Prognostic/Predictive Biomarker Read-Out. Cancers (Basel). 2021;13(23). **IF: 6.575**

Velásquez E*, **Szeitz B***, …, Domont GB, Rezeli M. Topological Dissection of Proteomic Changes Linked to the Limbic Stage of Alzheimer's Disease. Front Immunol. 2021;12:750665. **IF: 8.787**

**Shared first authors.*

Radeczky P, …, **Szeitz B**, …, Megyesfalvi Z, Dome B. Bone-Specific Metastasis Pattern of Advanced-Stage Lung Adenocarcinoma According to the Localization of the Primary Tumor. Pathol Oncol Res. 2021;27:1609926. **IF: 2.874**

Betancourt LH, Gil J, Kim Y, …, **Szeitz B**, …, Németh IB, Marko-Varga G. The human melanoma proteome atlas-Defining the molecular pathology. Clin Transl Med. 2021;11(7):e473. **IF: 8.554**

Betancourt LH, Gil J, …, **Szeitz B**, …, Marko-Varga G. The Human Melanoma Proteome Atlas-Complementing the melanoma transcriptome. Clin Transl Med. 2021;11(7):e451. **IF: 8.554**