DOI:10.14753/SE.2024.3031

SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3031.

SZEITZ BEÁTA

Diagnosztikus, digitális és molekuláris patológia
című program

Programvezető: Dr. Kiss András, egyetemi tanár

Témavezető: Dr. Szász Attila Marcell, tudományos főmunkatárs

konzulensek: Dr. Rezeli Melinda és Dr. Horvatovich Péter

DOI:10.14753/SE.2024.3031

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 Ph.D habil
Consultants:	Melinda Rezeli, Ph.D
	Péter Horvatovich, Ph.D
Official reviewers:	Sándor Spisák, Ph.D
	Katalin Dezső, MD Ph.D habil

Head of the Complex Examination Committee: Miklós Tóth, MD D.Sc

Members of the Complex Examination Committee: Lilla Turiák, Ph.D
Tibor Glasz, MD Ph.D

Budapest
2024

DOI:10.14753/SE.2024.3031

TABLE OF CONTENTS

DOI:10.14753/SE.2024.3031

DOI:10.14753/SE.2024.3031

LIST OF ABBREVIATIONS

adj.	adjusted
ALK	anaplastic lymphoma kinase
ALKi	anaplastic lymphoma kinase inhibitor
ANOVA	analysis of variance
avr.	average
BH	Benjamini-Hochberg
CCLE	Cancer Cell Line Encyclopedia
CM	culture media
CP	cell pellet
CPTAC	Clinical Proteomic Tumor Analysis Consortium
CV	coefficient of variation
DDA	data-dependent acquisition
DEA	differential expression analysis
DIA	data-independent acquisition
ECM	extracellular matrix
EMT	epithelial-mesenchymal transition
FC	fold change
FDA	Food and Drug Administration
FDR	false discovery rate
FFPE	formalin-fixed paraffin-embedded
FPKM	fragments per kilobase million
GDSC1	Genomics of Drug Sensitivity in Cancer 1
GDSC2	Genomics of Drug Sensitivity in Cancer 2
GSEA	gene set enrichment analysis
iBAQ	intensity-based absolute quantification
IC50	half maximal inhibitory concentration
IHC	immunohistochemistry
KEGG	Kyoto Encyclopedia of Genes and Genomes
LADC	lung adenocarcinoma
LC	liquid chromatography
LFQ	label-free quantification

LC-MS/MS	liquid chromatography-tandem mass spectrometry
M	mean
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSigDB	Molecular Signature Database
N/A	not available
NAT	normal adjacent tissue
NE	neuroendocrine
NES	normalized enrichment score
nr.	number
ns	not significant
NSCLC	non-small cell lung cancer
ORA	over-representation analysis
OXPHOS	oxidative phosphorylation
PCA	principal component analysis
pGSEA	pre-ranked gene set enrichment analysis
pROI	proteomic region of interest
PTM	post-translational modification
qPCR	quantitative polymerase chain reaction
Q3	third quartile
r	Pearson correlation coefficient
ROI	region of interest
RTK	receptor tyrosine kinase
SCLC	small cell lung cancer
SD	standard deviation
singscore	single-sample gene set score
ssGSEA	single-sample gene set enrichment analysis
TIL	tumor infiltrating lymphocyte
tROI	transcriptomic region of interest

DOI:10.14753/SE.2024.3031

1. INTRODUCTION

According to the World Health Organization, lung cancer was the second most commonly diagnosed cancer in 2020 with 2.21 million cases and was the leading cause for cancer-related deaths with 1.8 million people dying of lung cancer (1, 2). High incidence and mortality rates of this cancer type are characteristic of Hungary as well (3). Lung cancer is a heterogeneous disease both in terms of its clinicopathological and molecular features, which poses several challenges in the diagnosis and treatment of patients (4).

Proteomics plays a central role in advancing our understanding of cancer biology. It refers to the large-scale investigation of proteins, their expression, structure, and physiological roles within biological systems. By quantifying the variations and abundance of proteins, we gain a deeper understanding of the function of individual proteins, and of their spatiotemporal dynamics within those systems (5). The strength of proteomics lies in its closeness to the phenotype; thus, it can complement (epi)genomics- and transcriptomics-based efforts in biomarker discovery and in capturing the complexity of the disease (6-10).

My doctoral thesis focuses on the heterogeneity of two lung cancer types, small cell lung cancers (SCLCs) and anaplastic lymphoma kinase (*ALK*)-rearranged lung adenocarcinomas (LADCs), which we investigated using mass spectrometry (MS)-based proteomics and complemented by transcriptome-based findings. The thesis thus continues with an introduction to these lung cancer types, followed by an overview of MS-based proteomics.

1.1 Lung cancer

Histologically, two main types of lung cancer exist: SCLC (representing ca. 15% of all cases) and non-small cell lung cancer (NSCLC, accounting for ca. 85% of the cases). Within the NSCLC histology type, adenocarcinomas are the most prevalent, followed by squamous-cell carcinomas (11).

1.1.1 Small cell lung cancer

While less frequent, SCLC is distinguished by its aggressive behavior and early metastatic spread, resulting in a 5-year survival rate of less than 7% (12). Histologically, SCLC tumor cells are of small size with round-to-fusiform shape, their cytoplasm is generally scanty, have finely granular nuclear chromatin and either absent or inconspicuous nucleoli (13). At the molecular level, these tumors are characterized by high mutation burden, extensive chromosomal rearrangements, and in general, the tumor suppressors *TP53* and *RB1* are functionally inactivated (14). Oftentimes, diagnosis occurs when patients are in the extensive stage of the disease, making surgical intervention no longer feasible and limiting treatment options to cytotoxic chemotherapy and radiation (14). Immune checkpoint inhibitors have been administered only with moderate success to SCLC patients, and targeted therapies have so far failed, which emphasizes the importance of basic discovery and clinical translational research to develop more effective therapeutic approaches for SCLC (15).

Recent molecular profiling studies of SCLC shed light on substantial heterogeneity, introducing a new classification system based on both neuroendocrine (NE) features and unique molecular profiles (12, 16, 17). Based on the expression patterns of transcription factors *ASCL1*, *NEUROD1*, *POU2F3* and *YAP1*, Rudin et al. proposed new molecular subgroups, namely SCLC-A, SCLC-N, SCLC-P and SCLC-Y. These subtypes show diverse NE characteristics, with SCLC-A and SCLC-N being classified as NE tumors (NE high and low, respectively), while SCLC-P and SCLC-Y were described as non-NE tumors (16). The classification system was established based on genomic, epigenetic, and transcriptomic profiling studies performed on clinical samples and preclinical models (16, 18-20). However, follow-up immunohistochemistry (IHC) analyses of human tumor tissue samples demonstrated that not all relationships between subtype markers and NE features, as initially anticipated from gene expression data, could be confirmed by IHC (21). This study also failed to validate a unique *YAP1*-driven subtype (21). An inflamed subtype (SCLC-I) characterized by an inflamed gene signature was proposed as an alternative to the *YAP1*-defined subtype (22). This underpins the importance of protein-level investigations to better understand the molecular subtypes of SCLC and to identify diagnostic markers as well as the subgroups' therapeutic vulnerabilities.

However, as of today, SCLC heterogeneity is understudied at the protein level. The first large-scale proteogenomic analysis of 112 SCLC tissues was only published this year (23). Previous proteomic studies treated SCLC as a single entity and SCLC tumors were compared with normal bronchial epithelial tissues (24), non-SCLC cell lines (25), carcinoid tumor tissues (26), or the cell lines' proteomic profile was published as part of a larger cell line proteome resource (27).

1.1.2 *ALK*-rearranged lung adenocarcinoma

LADC patients have a dismal five-year survival rate of 15%, attributable to the fact that diagnosis usually occurs at a late stage and patients develop resistance to treatment (28). Multiple morphological subtypes of LADC exist, including acinar, lepidic, (micro)papillary and solid morphologies (29). For example, the acinar (also called tubular) morphology refers to glandular structures with luminal spaces. Papillary morphology is described as finger-like projections lined by tumor cells. In contrast, sheets or nests of tumor cells displaying no glandular or papillary structures is characteristic of the solid subtype (30). Comprehensive histopathological assessment of lung cancers entails the description of various tumor features besides morphology, such as the density and distribution of tumor infiltrating lymphocytes, stroma, and mucin production. Alongside with the prognostic and therapeutic relevance of tumor morphology (30), these histopathological features influence tumor behavior, potentially impacting therapy resistance and patient outcome (31-33).

Considering molecular heterogeneity, numerous genetic aberrations may drive lung carcinogenesis, as revealed by whole-genome/-exome and RNA-sequencing studies. These aberrations include *EGFR*, *KRAS*, *BRAF* and *HER2* mutations and gene rearrangements involving *ALK*, *RET*, *ROS1* and skipping of exon 14 in *MET*. Currently, only targeted therapies against *EGFR* mutations and *ALK* fusions are part of standard care next to conventional therapies (surgery, chemo- and radiotherapy) and immunotherapy (34).

Merely 3-5% of LADCs from the Caucasian population harbor *ALK* gene rearrangements (35). Upon fusion with partner genes (most frequently *EML4*), *ALK* activity increases, activating downstream signaling pathways including MAPK, PI3K-AKT, or JAK-STAT (36). This promotes the proliferation and survival of cancer cells.

The aberrantly activated *ALK* can be targeted with ALK inhibitors (ALKi) such Crizotinib (Xalkori; first-generation ALKi) or Alectinib (Alecensa; second-generation ALKi). While these therapies significantly improve patient outcomes, most patients experience tumor recurrence due to cancer cells acquiring resistance to treatment (37). Intratumoral heterogeneity is a crucial contributor to therapeutic resistance. It is a complex phenomenon influenced by genetic, epigenetic, and environmental factors, and is affected by the interplay between the tumor cells and their microenvironment (38). Notably, lung cancers driven by *ALK* rearrangements often contain *ALK*-negative regions as well (39), potentially introducing additional molecular- and pathway-level heterogeneity within these tumors.

Large cohorts of LADCs were previously investigated via multiple omic approaches including proteomics (6-8, 10). However, these cohorts contain only a few *ALK*-driven tumors, and bulk-tissue expression profiling was performed. In parallel, spatially resolved profiling studies on non-*ALK*-driven NSCLCs demonstrated the value of the multiplex and targeted RNA/protein analyses for biomarker discovery (40-43). This highlights the unmet need to map the multi-omic profiles of *ALK*-rearranged LADCs while also capturing their intratumoral heterogeneity.

1.2 Mass spectrometry-based proteomics

Mass spectrometry (MS)-based proteomics has become the primary tool to explore entire proteomes in-depth, including the analysis of PTMs and cancer-specific protein sequence alterations (44).

The value in the direct investigation of the entire proteome stems from the intricate and dynamic nature of proteome regulation. Protein expression is influenced by factors such as alternative splicing, single nucleotide polymorphisms (resulting in diverse proteoforms) and transcript degradation. In addition, various processes taking place at proteome level, including protein-protein interactions, degradation rates and post-translational modifications (PTMs), contribute to the intricate nature of the proteome (45, 46). In line with this, multi-omic cancer studies in the past demonstrated only a weakly positive agreement between the samples' mRNA and protein expression profiles (47-50), which emphasizes the importance of directly mapping the cancer proteome.

There are two main techniques for the MS-based analysis of proteins, called the "top-down" and "bottom-up" approaches. In the top-down methodology, proteins are analyzed in their intact form. This requires a more sophisticated instrumentation, and only a limited number of proteins can be analyzed at the same time (51). In contrast, the bottom-up methodology (also called shotgun proteomics) tackles the analysis of the proteome by first subjecting proteins to enzymatic digestion, and then by analyzing the resulting peptides. Thus, proteins are inferred based on the measured peptides (52). The bottom-up approach is a well-established technique, enabling the identification and quantification of thousands of proteins in a single sample (53), thus suitable, for instance, in the discovery phase of cancer studies to provide an in-depth view of the tumors' proteome (54). In further paragraphs of the *Introduction*, only the bottom-up approach in the context of discovery proteomic studies is discussed.

Proteomic experiments start with sample processing, which is instrumental in eliminating inferring compounds, in maximizing the number of peptide identifications, and in ensuring reproducibility. Tumors may undergo tissue sectioning and histological examination prior to proteomic sample preparation. Exact steps, such as protein extraction, digestion (oftentimes with trypsin), peptide fractionation, labeling approaches for protein quantification, or enrichment of peptides with specific PTMs, are decided based on the clinical question and are also determined by the tumors' storage method (fresh-frozen or formalin-fixed paraffin-embedded, FFPE). Regardless of the exact sample preparation workflow, complex peptide mixtures are ultimately injected onto a liquid chromatography (LC) system coupled with electrospray ionization tandem MS (MS/MS). Peptide ions are first separated on a reversed-phase chromatographic column, followed by mass measurement (MS1), and then are fragmented and subjected to a second stage of mass measurement (MS2) (55). MS data acquisition can be data-dependent and data-independent acquisition (DDA and DIA, respectively). DDA involves acquiring an MS1 spectrum followed by selecting a set number of precursor peptide ions for fragmentation. In contrast, DIA acquires MS2 spectra without bias, sequentially fragmenting all peptides in a specified mass-to-charge ratio range. While DDA has straightforward mass spectra interpretation, its semi-stochastic nature leads to low reproducibility and may overlook less abundant precursor ions. DIA offers precise quantitation, increased protein identification, but poses computational challenges and

complexities in raw data interpretation (56). Collectively, the gathered liquid chromatography-tandem mass spectrometry (LC-MS/MS) data can be used to both identify and quantify peptides in a sample with dedicated software (55).

1.3 Proteomic data processing and analysis

The raw data obtained from LC-MS/MS require sophisticated algorithms for peptide identification, quantification, and protein inference. Commonly used raw data processing tools include Proteome Discoverer (Thermo Scientific), FragPipe (57), MaxQuant (58), Skyline (59), MetaMorpheus (60), and Spectronaut (Biognosis).

Peptide identification algorithms are categorized broadly based on how the search space is constructed, i.e. the list of sequences that are considered when matching spectra to peptides. The *de novo* search considers all possible amino acid sequences (61), the database search uses *in silico*-digested protein sequences which protein sequences are expected to be present in the sample (e.g. human protein sequences for human samples) (62), and finally, the spectral library search matches the collected MS2 spectra with entries in a reference library, hence the search space is limited to peptides previously identified via MS (63). After peptide identifications were made by the algorithms, the potential peptide-spectrum matches are subjected to a subsequent statistical validation step to discriminate between correct and incorrect identifications. The false discovery rate (FDR) is often computed via the target-decoy strategy, in which case the decoy hits are used to estimate a threshold for retaining high-confidence target peptide-spectrum matches (64). Protein inference takes place after peptide identification, which includes grouping peptide-spectrum matches according to their corresponding protein and employing another statistical data validation step at the protein level (64). For abundance estimation of proteins, label-free quantification (LFQ) and label-based quantification approaches were developed. LFQ relies either on spectral counting, or on quantitation based on the precursor ions' chromatographic peak areas. Labeling is based on subjecting peptides to metabolic and chemical modifications, thus requiring extra steps during sample preparation but providing the highest precision in quantification. In contrast, LFQ is cost-efficient from a sample preparation point-of-view and still results in comparable proteome coverage (65).

Ultimately, data processing software generates a protein abundance matrix listing proteins and their abundances per sample, which is then subjected to post-processing, followed by subsequent downstream (statistical and bioinformatic) analyses (Figure 1).

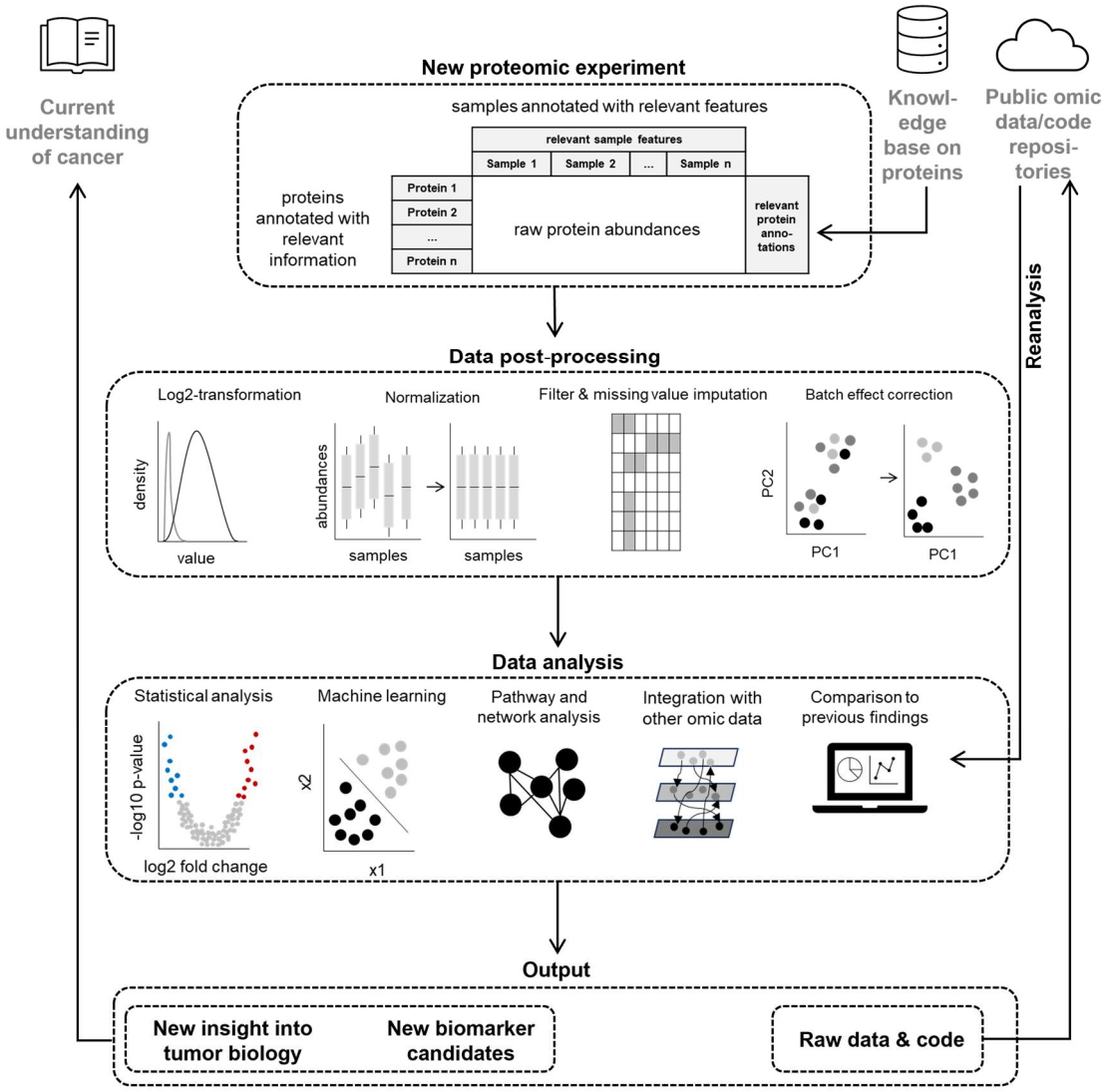

Figure 1. A schematic data analysis workflow in discovery proteomic studies.

Post-processing includes log2-transformation of protein abundances to remove the extreme skewness of the values towards zero and to stabilize the variability of protein abundances (66). Run-to-run variability between measurements, stemming from non-biological factors, needs to be mitigated with normalization, e.g. via using the median, linear regression, variance stabilization, or quantile normalization strategies (67). Given that normalization only aligns the samples' global patterns, further batch effects,

stemming from differences in sample preparation and data acquisition conditions, might still be present and need to be corrected for (68). In addition, proteins quantified only in relatively few samples are removed prior to an optional missing data imputation step, and before statistical analyses. The imputation method is chosen based on whether the missingness is assumed to be intensity-independent or -dependent (69). Popular methods include k-nearest neighbors (70) (suited for intensity-independent missingness), imputation of values from a normal distribution (71) and imputation by accelerated failure model (72) (both suited for intensity-dependent missingness).

After post-processing, the protein abundance matrix together with the cleaned sample metadata, is ready for subsequent downstream analyses. The data analysis steps are chosen based on the original research question. Many studies aim to understand how proteomic profiles relate to specific phenotypes (73). These can be answered with traditional statistical tests such as t-tests, analysis of variance (ANOVA) tests, or linear regression. To address the multiple hypothesis testing problem that arises with such investigations, p-value adjustment is needed to effectively reduce the number of false positive findings, e.g. via the Bonferroni, Holm, or the Benjamini-Hochberg (BH, also known as FDR) procedures (74). In addition, machine learning-based approaches such as unsupervised clustering, principal component analysis (PCA), partial least squares-discriminant analysis or support vector machine algorithms are used to explore the high-dimensional proteomic data at hand (73).

Additional knowledge on protein function, cellular localization, protein-protein interactions and on regulatory processes, can be incorporated in proteomic data analysis by approaches termed as “pathway analysis” and “network analysis”. These analyses rely on knowledge bases like Gene Ontology, Reactome, Kyoto Encyclopedia of Genes and Genomes (KEGG), Molecular Signature Database (MSigDB), STRING, or BioCarta. Network analysis builds protein interaction networks using previous experimental data and new *in silico* predictions to gain systems-level biological understanding. In contrast, pathway analysis identifies pathways that may explain the molecular mechanisms that resulted in the proteins’ altered presence or their differential abundance between sample phenotypes (75). Popular methods include over-representation analysis (ORA), which assesses the enrichment of a predefined gene set within a subset of “interesting” proteins, comparing its presence to what would be expected by chance (76). Gene set enrichment

analysis (GSEA) is a rank-based approach that assesses whether predefined groups of proteins are predominantly upregulated or downregulated in one sample phenotype compared to another. GSEA is particularly valuable when the differential expression analysis falls short in detecting subtle yet coordinated changes in expression across sample phenotypes for groups of related proteins (77).

Proteomics, whether used alone or integrated with other omic data, offers valuable insights into tumor biology, and aids the identification of diagnostic, predictive, prognostic, and therapeutic markers. To promote reproducibility, as well as validation and meta-analysis by subsequent research endeavors, it is desirable to share the data and code with the scientific community (78). Similarly, public omic data repositories can be utilized to validate study findings by reanalyzing prior independent research.

DOI:10.14753/SE.2024.3031

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.

DOI:10.14753/SE.2024.3031

3. METHODS

3.1 Characteristics of studied biological specimens

3.1.1 Small cell lung cancer cell lines

The 26 human SCLC cell lines were purchased from American Type Culture Collection or were kindly provided by our collaborators (Table 1) (79).

Table 1. Cell lines subjected to proteomics, and their general characteristics. Abbreviations: N/A, not available; SCLC, small cell lung cancer. Modified table from (79).

Cell line ID	Subtype	Cell line origin	Chemotherapy	Culture type
DMS153	SCLC-A	metastatic	post-chemo	semi-adherent
DMS53	SCLC-A	lung	chemo-naïve	adherent
H146	SCLC-A	metastatic	chemo-naïve	suspension
H1688	SCLC-A	metastatic	chemo-naïve	adherent
H1882	SCLC-A	metastatic	N/A	adherent
H209	SCLC-A	metastatic	chemo-naïve	suspension
H378	SCLC-A	lung	post-chemo	suspension
SHP77	SCLC-A	lung	N/A	adherent
GLC4	SCLC-N	pleural effusion	chemo-naïve	suspension
H1694	SCLC-N	lung	N/A	semi-adherent
H2171	SCLC-N	pleural effusion	post-chemo	suspension
H446	SCLC-N	pleural effusion	N/A	adherent
H524	SCLC-N	metastatic	post-chemo	suspension
H82	SCLC-N	metastatic	N/A	semi-adherent
N417	SCLC-N	lung	N/A	suspension
COR-L311	SCLC-P	lung	post-chemo	suspension
H1048	SCLC-P	pleural effusion	N/A	adherent
H211	SCLC-P	lung	post-chemo	suspension
H526	SCLC-P	metastatic	chemo-naïve	suspension
CRL-2066	SCLC-Y	lung	chemo-naïve	adherent
CRL-2177	SCLC-Y	lung	N/A	adherent
H1341	SCLC-Y	metastatic	N/A	adherent
H196	SCLC-Y	pleural effusion	post-chemo	adherent
H372	SCLC-Y	metastatic	N/A	adherent
H841	SCLC-Y	lung	post-chemo	adherent
HLHE	SCLC-Y	metastatic	N/A	adherent

The cell lines were maintained at 37°C in a humidified incubator with 5% CO2, in RPMI-1640 with 10% fetal calf serum (Sigma Chemical Co.), 100 U/ml penicillin and 10 mg/ml streptomycin (Sigma Chemical Co.) Importantly, cell lines were regularly checked for mycoplasma contamination and were used within 10 passages after authentication. Subtype (SCLC-A/N/P/Y) was determined by analyzing the *ASCL1*, *NEUROD1*, *POU2F3* and *YAP1* gene expression patterns via quantitative polymerase chain reaction (qPCR) (79).

3.1.2 *ALK*-rearranged LADC tissues

The primary LADC tumors were obtained from the National Korányi Institute of Pulmonology in Hungary. IHC and fluorescence in situ hybridization were used to determine *ALK* positivity at the 2nd Department of Pathology, Semmelweis University, Hungary. Ethical approval was granted by the Medical Research Council of Hungary (2521-0/2010-1018EKU, 510/2013, 52614-4-213EKU). All patients were treated with ALK inhibitors after determining *ALK* positivity (80).

The proteomic and transcriptomic regions of interest (pROIs and tROIs, respectively) underwent histopathological evaluation, for which the slides were stained with hematoxylin and eosin and scanned with a Pannoramic 250b Slide scanner (3DHistech Ltd.). Both tumor and normal adjacent tissue (NAT) regions were selected for molecular profiling. The morphological areas were manually annotated. Immune scores (spanning from 0 to 3) were derived from a semi-quantitative assessment of immune cell (lymphocytic) infiltration within the area of tumor cell nests (80). This was performed using the QuPATH v0.3.0 software (81). Mucin and stroma scores were given as follows: score 0: 0%, score 1: 1-33%, score 2: 34-66%, score 3: >66% mucin or stromal content in the given ROI. The clinical and histopathological data of this cohort are summarized in Table 2 (80).

Table 2. Summary of clinical data and the histopathological data of ROIs. Abbreviations: ALKi, ALK inhibitor; avr., average; nr., number; N/A, not available; NAT, normal adjacent tissue; pROI, proteomic region of interest; TIL, tumor infiltrating lymphocyte; tROI, transcriptomic region of interest; yrs, years. Modified table from (80).

	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**	**Case 7**
Sex	male	male	male	female	female	female	male
Age at diagnosis (years)	53.6	43.7	68.9	32.8	68.5	64.2	66.7
Stage on presentation	N/A	N/A	N/A	3	N/A	4	3
ALK inhibitor	Crizo-tinib	Crizo-tinib	Crizo-tinib	Crizo-tinib	Crizo-tinib	Alec-tinib	Alec-tinib
Alive	no	yes	no	yes	yes	yes	yes
Overall survival (years)	2.2	6.6	4.1	13.8	7.3	4.9	6.7
pROI morphology (nr. of pROIs)	tubular (2)	NAT (1), solid (1)	NAT (3), papillary (2), tubular (1)	papillary (3)	solid (4)	solid (3)	NAT (1), solid (2)
Avr. TIL of pROIs (%)	25.00	25.00	15.00	23.33	8.75	28.33	30.00
Avr. mucin score of pROIs	3.00	2.00	2.00	3.00	0.00	0.00	0.50
Avr. stroma score of pROIs	3.00	2.00	1.67	2.33	1.75	1.00	3.00
tROI morphology (nr. of tROIs)	NAT (1), tubular (11)	NAT (2), solid (10)	NAT (2), papillary (8), tubular (2)	NAT (1), papillary (11)	solid (12)	solid (12)	NAT (2), solid (10)
Avr. immune score of pROIs	2.25	2.00	2.00	1.42	2.25	1.60	1.17

3.2 Experimentally collected proteomic and transcriptomic data

3.2.1 Proteomic analysis of SCLC cell lines

Both the pellets and media (CP and CM, respectively) from 26 cell lines were processed and subjected to MS-based proteomic analysis.

In brief, the CPs were solubilized with a protein extraction buffer containing 25 mM dithiothreitol, 10% sodium dodecyl sulphate, 100 mM triethylammonium bicarbonate (pH = 8). The volume of the buffer was determined by the cell count, where

250 µl of protein extraction buffer was added to samples containing 5 million cells. Solubilization was achieved by incubation for 5 min at 95°C with shaking at 500 rpm. Protein extraction was carried out through a 20 min sonication at 4°C (Bioruptor Plus, Diagenode) with 40 cycles (15 s on/15 s off), followed by a brief centrifugation at 20000 × g at 18°C to remove cell debris. Protein concentration was determined using a Pierce 660 nm Protein Assay kit (Thermo Scientific) (79).

The filtered CM samples underwent concentration using spin concentrators (5K 4 ml, Agilent Technologies) until reaching ca. 100 µl in volume. NanoDrop (DeNovix DS-11 FX +) was used for protein determination. Sodium dodecyl sulfate was added to a final concentration of 3%, as well as 100 mM triethylammonium bicarbonate was added to adjust the pH. Reduction was carried out using 10 mM dithiothreitol with a 1-hour incubation period at 37°C (79).

For both CP and CM, protein digestion was achieved utilizing the S-Trap technology (ProtiFi) with slight modifications, as previously described by our group (82). In short, the samples underwent alkylation with 50 mM iodoacetamide and acidification with 1.2% phosphoric acid. Subsequently, S-Trap binding buffer (90% methanol, 100 mM triethylammonium bicarbonate) was added to 7× the final sample volume, and the mixtures were transferred to the S-Trap 96-well digestion plate. Captured proteins were subjected to four washes with 200 µl of S-Trap binding buffer followed by brief centrifugations (2 min at 1000 × g). For digestion, a buffer containing 50 mM triethylammonium bicarbonate along with LysC at a 1:50 enzyme-to-protein ratio was added on top of the filters. The samples were incubated for 2 h at 37°C. Next, a digestion buffer containing trypsin at a 1:50 enzyme-to-protein ratio was added, and the samples were incubated overnight at 37°C. The next day, peptides were eluted in three steps: initially with 80 µl of digestion buffer, followed by 80 µl of 0.2% formic acid, and finally with 80 µl of 50% acetonitrile containing 0.2% formic acid. Peptides were dried down via a vacuum concentrator. The Pierce Quantitative Colorimetric Peptide Assay kit (Thermo Scientific) was used for peptide determination (79).

A Q Exactive HF-X mass spectrometer coupled to a Dionex UltiMate 3000 RSLCnano UPLC system (Thermo Scientific), equipped with an EASY-Spray ion source was used for the nano LC-MS/MS analysis. Peptides from CP and CM samples were injected in triplicate (1.5 µg and 1 µg respectively). Samples were loaded onto an Acclaim

PepMap 100 C18 trap column (75 µm × 2 cm, 3 µm, 100 Å, nanoViper) and were separated on an Acclaim PepMap RSLC C18 column (75 µm × 50 cm, 2 µm, 100 Å) (Thermo Scientific). A flow rate of 300 nl/min, a column temperature of 60°C, and a 145-minute-long gradient was applied using solvents A (0.1% formic acid) and B (0.1% formic acid in 80% acetonitrile). Solvent B was increased from 2% to 25% over 115 min, then to 32% in the next 10 min, and further to 45% in 7 min. Lastly, solvent B was raised to 90% in 8 min, then this concentration was maintained for an additional 5 min (79).

Considering the MS approach, peptides from the CP were analyzed with one DDA and two DIA runs, while peptides from the CM were subjected to two DDA and one DIA runs. For the top 20 DDA method, full MS1 scans were conducted at m/z 375-1500, resolution of 120000 (at 200 m/z), target automatic gain control value of 3×10^6 and maximum injection time of 100 ms. A normalized collision energy of 28 was used for fragmentation, with an isolation window set to 1.2 m/z. MS2 scans were acquired at a resolution of 15000 (at 200 m/z), with a target automatic gain control value of 1×10^5, a maximum injection time of 50 ms, an ion selection threshold of 8×10^3, and dynamic exclusion set to 40 s. Regarding the DIA analysis, a complete acquisition cycle comprised three MS1 full scans, each followed by 18 MS2 DIA scans with variable isolation windows. MS1 full scans were conducted within the range of m/z 375-1455, with a resolution of 120000 (at 200 m/z), a target automatic gain control value of 3×10^6, and a maximum injection time of 50 ms. MS2 scans were acquired with a resolution of 30000 (at 200 m/z), using a normalized collision energy of 28 for fragmentation, a target automatic gain control value of 1×10^6, automatic maximum injection time, and a fixed first mass of 200 m/z. The used variable isolation windows were 13.0, 16.0, 26.0, and 61.0 m/z, with 27, 13, 8 and 6 loop counts, respectively (79).

The database search was performed within Proteome Discoverer v2.4, using the SEQUEST HT search engine combined with spectral library search. The UniProtKB human database (accessed on 15 January 2019) and the Proteome Tools spectral libraries were utilized during the search. Carbamidomethylation of cysteine was set as static modification, while oxidation of methionine and N-terminal acetylation were included as dynamic modifications. A precursor tolerance of 10 ppm and a fragment mass tolerance of 0.02 Da were applied. Up to two missed cleavages, and an FDR of 1% at the peptide

and protein levels were allowed. The top three average method was employed for protein quantitation (79).

3.2.2 Proteomic analysis of LADC tumors with *ALK*-rearrangement

The sample preparation steps and MS measurements for the proteomic data used in our study was initially reported in (83), following previously established protocols described in (84, 85). The tissue areas selected for proteomic analysis can be seen in Figure 2. In short, the sample preparation included dewaxing, rehydration, and antigen retrieval of the selected ROIs, followed by on-surface digestion with a Trypsin/Lys-C mix, peptide extraction and purification. Reversed-phase peptide separation was performed on an Acquity M-Class BEH130 C18 analytical column using Dionex Ultimate 3000 RSLC nanoUHPLC. MS analysis in DDA mode was employed using a Bruker Maxis II Q-TOF (Bruker Daltonik GmbH) MS. Proteins were identified with Byonic v4.2.10 (Swiss-Prot human database, accessed on 01 November 2020), and a focused database was created for subsequent label-free quantitation with MaxQuant v1.6.17 (80, 83).

3.2.3 NanoString GeoMx profiling of LADC tumors with *ALK*-rearrangement

For the NanoString GeoMx profiling, FFPE slides were deparaffinized and rehydrated, followed by antigen retrieval and digestion by proteinase-K. Following an overnight hybridization with cancer transcriptome atlas probes, the samples were washed to remove off-target probes and stained with morphology markers (PanCK, SYTO83, CD45, CD3). From each ROI, RNA identification and unique molecular identifier-containing oligonucleotide tags were cleaved with ultraviolet light and collected by the GeoMx. The digital slides with this fluorescent staining were inspected, and ROIs were selected (Figure 2). Next-generation sequencing library preparation and purification was performed after collecting the oligo tags, and sequencing was carried out on an Illumina NovaSeq 6000 (Illumina). The Nanostring DnD pipeline was utilized to process fastq files into digital count (.dcc) files, which were uploaded into the GeoMx software for data post-processing (80).

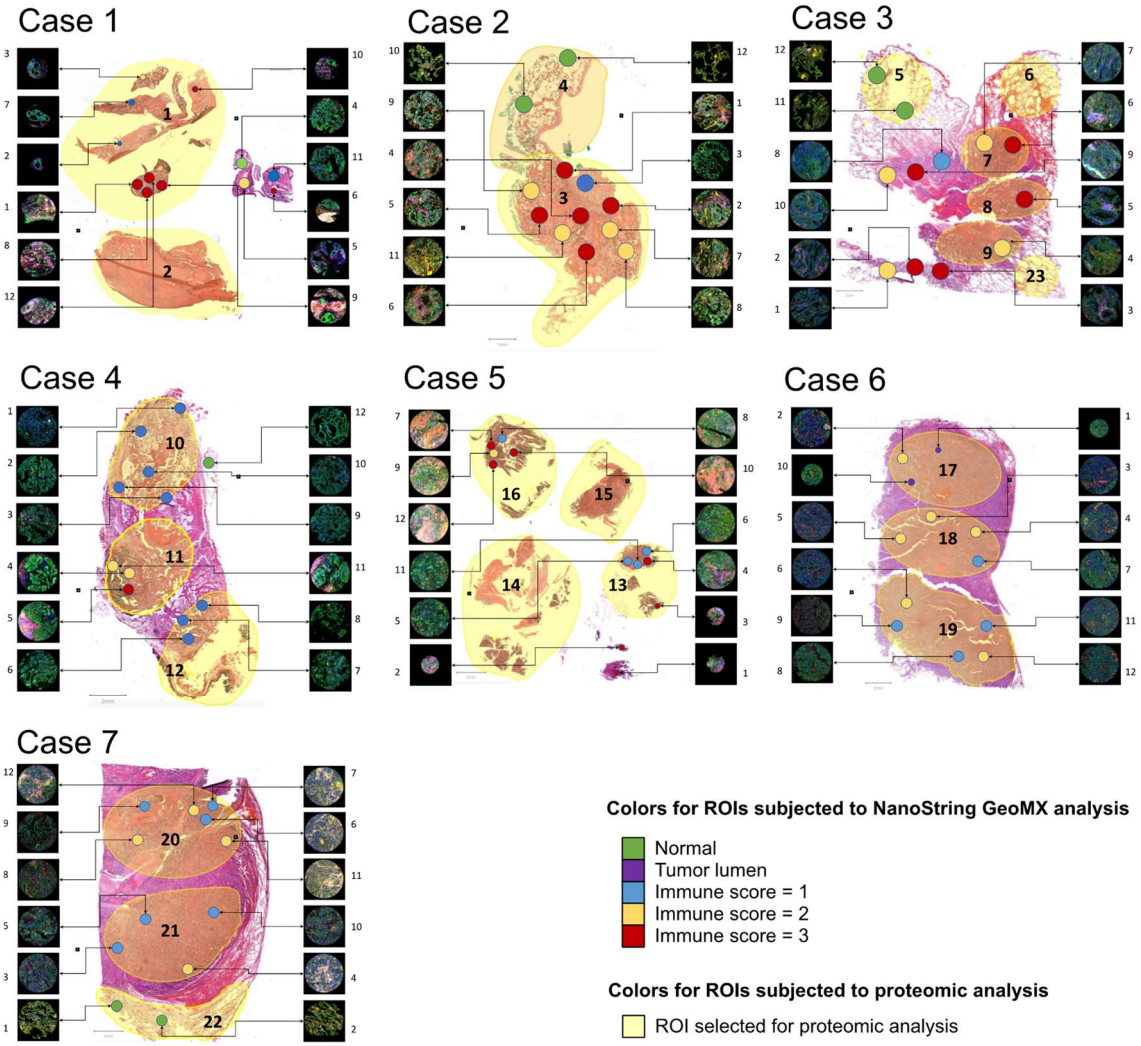

Figure 2. The analyzed regions across the seven *ALK*-rearranged lung adenocarcinomas. The smaller transcriptomic regions of interest (tROIs) are colored based on immune score or whether the tROI was in a normal adjacent tissue (NAT) region or at the site of vascular invasion ("tumor lumen"). Regions for proteomic analysis covered larger tumor and NAT areas. Hematoxylin and eosin slides: 0.4x magnification, fluorescent staining inserts: 63x magnification. Figures in a modified form from (80).

3.3 Data cleaning and post-processing

Data analysis steps were mainly performed with R v4.2.0 (R Foundation for Statistical Computing, Vienna, Austria, 2022) and the custom R scripts were uploaded to GitHub: https://github.com/bszeitz/SCLC_proteomics (*study I*) and https://github.com/bszeitz/ALK_rearranged_pADCs_multiomics (*study II*).

3.3.1 Proteomic data post-processing

Similar post-processing steps were conducted in *studies I* and *II*. The protein intensities (LFQ or intensity-based absolute quantification (iBAQ) values) were log2-transformed and normalized by centering the values around the global median in each sample (79, 80). Further post-processing steps were done in *study I* as the same cell lines were subjected to LC-MS/MS analysis multiple times. Firstly, triplicate measurements from the same MS vial were checked for low-quality measurements, and those were subsequently excluded. The median value of protein intensity measurements from the same MS vial was then calculated. The batch effect was removed using univariate linear regression, after which all replicates from the same cell line could be averaged (79).

The filter for proteins with quantitative values across at least 80% of the samples was performed both in *studies I* and *II*. In addition, "on/off" proteins across SCLC subtypes were identified in *study I* based on their missing value patterns. For this analysis, proteins present in ≥85% of samples in one SCLC subtype and ≤15% in other subtypes were considered "on", while proteins present in ≤15% of samples in one subtype and ≥85% in other subtypes were considered "off" in a certain subtype. To impute missing values with low-intensity values, the Perseus v1.625 (71) software was used in *study I* (impute low-intensity values based on normal distribution, width = 0.3, down shift = 1.8), whereas in *study II*, the "impute.MinProb" function from the imputeLCMD R package was used (79, 80).

3.3.2 NanoString data post-processing

The raw NanoString data were first processed in the GeoMX Analysis Software v2.2.0.122. All tROIs passed initial quality control, which included checking the sequencing parameters and the template control counts. Biological Probe quality control was then run to identify any outlying probes (a total of five probes per gene target) before individual probe data were collapsed to gene-level counts. A total of nine outlying gene probes from nine separate genes, as well as one outlying negative probe was detected and removed. In the resulting raw, gene-level count data, the negative probes were used to calculate the limits of quantitation for each tROI using the following formula: [geometric mean of negative probes × (geometric standard deviation of negative probes)^2]. For third quartile (Q3) normalization of the raw counts, a normalization factor was generated for

each tROI as shown: [geometric mean of all ROI Q3 values in the dataset / individual ROI Q3 value]. The final data contained the normalized gene counts of 1811 genes, quantified across all tROIs (80).

3.3.3 Accessing external databases and datasets

To annotate proteins from *study I*, various databases were utilized, including databases on secreted proteins (86-88), on proteins detectable/actively secreted in the blood (the Human Protein Atlas v21.1, https://www.proteinatlas.org), on "druggable" proteins (the Human Protein Atlas v21.1). Functional proteins annotations were extracted from UniProt, Release 2022_03 (89). The list of Food and Drug Administration (FDA)-approved drugs that directly interact with selected proteins as part of their mechanism of action was gathered from the DrugBank database (90). In addition, the CancerRxGene drug sensitivity data [Release 8.3 (91); Genomics of Drug Sensitivity in Cancer 1 (GDSC1) and GDSC2 datasets] was downloaded from the FTP Server of the Wellcome Sanger Institute (79).

For *study I*, the transcriptomic data (both the Fragments per kilobase million (FPKM) and the Z-score values) of relevant SCLC tissue samples were accessed from cBioPortal (92, 93), downloading George et al.'s work (18). In addition, the raw RNA-Seq data of Cancer Cell Line Encyclopedia (CCLE) cell lines (gene count data, normalized using RNA-Seq by Expectation-Maximization, i.e. RSEM method) were accessed from the Cancer Dependency Map (94). This raw data were further processed using the limma R package (95), which included normalization factor calculation, filtering out genes with only expression values below 50, and voom transformation where the model matrix included the subtype assignment (79).

To supplement results in *study II*, the LADC proteomic and RNA-Seq dataset of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) (7) was accessed (80).

3.4 Data analysis

3.4.1 Calculation of sample-wise scores

In *study I*, the transcriptomic dataset from George et al. (18) was subjected to single-sample gene set enrichment analysis (ssGSEA) (96). Only transcripts with sum FPKM higher than 50, and one transcript per gene (the transcript with the highest sum

FPKM value) was retained. The normalized enrichment score (NES) per sample was calculated for selected gene sets with the following settings: sample.norm.type = “rank”, weight = 0.75, statistic = “area.under.RES”, output.score.type = “NES”, nperm = 1000, min.overlap = 5, and correl.type = “z.score”. To quantify the NE and epithelial-mesenchymal transition (EMT) characteristics of SCLC cell lines, previously published gene sets (Zhang et al. (17) for NE markers, Kohn et al. (97) for epithelial and mesenchymal markers) were leveraged. The NE score was calculated for each cell line in the following manner: NE score = (Mean Z-scored protein expression of NE markers) - (Mean Z-scored protein expression of non-NE markers). Similarly, the EMT score was calculated the following way for each cell line: EMT score = (Mean Z-scored protein expression of mesenchymal markers) - (Mean Z-scored protein expression of epithelial markers) (79).

In *study II*, the normalized gene counts from NanoString and the iBAQ protein intensities after missing value imputation were transformed into single-sample gene set scores (singscores) using the singScore R package (98). The Hallmark (99), KEGG (100) and Reactome (101) gene sets were obtained from MSigDB v7.5.1 (102) for this analysis. Minimum 10 overlapping genes between our gene/protein expression data and the gene set were expected prior to the singscore calculation, other parameters were left at default (80).

3.4.2 Cluster analysis

Clustering settings for heat map visualizations via the ComplexHeatmap R package (103) were hierarchical clustering, Euclidean distance and complete linkage, unless specified otherwise. To discover sample groups in an unsupervised manner, a consensus clustering algorithm implemented in the ConsensusClusterPlus R package (104) was used in both studies. In *study I*, the basis of clustering was the samples’ protein expression profile in the CP dataset after filtering for proteins that have a standard deviation (SD) above 1.25. In *study II*, the pROIs’ protein LFQ values and tROIs’ gene counts after Z-score normalization were used. Data were resampled 1000 times via the bootstrap method with a probability of 0.8 for selecting any item (i.e. sample) and any feature (i.e. protein/gene), then these bootstrap sample datasets were clustered using the partitioning around medoids method with Pearson distance and complete linkage. The

best number of clusters was selected based on the visual inspection of ConsensusClusterPlus outputs in both studies, plus considering silhouette width information in *study I* (79, 80).

3.4.3 Correlation analysis

Before assessing the similarity of the tROIs and pROIs in *study II*, the gene counts and singscores of the tROIs from the same pROI were averaged. These averaged values were then correlated with the pROIs' LFQ values and singscores (Pearson correlation test, via the correlation R package). BH-adjusted (adj.) $p < 0.05$ was considered significant (80).

3.4.4 Differential expression analysis

In *study I*, ANOVA was used to perform differential expression analysis (DEA), followed by Tukey's honestly significant difference post hoc tests. A protein was considered significant if both the BH-adj. p-value of the ANOVA test and the relevant pairwise Tukey's honestly significant difference test p-values were less than 0.05. For the SCLC transcriptomic data from the CCLE database (94), limma was used for DEA by fitting a linear model, followed by Empirical Bayes smoothing of standard errors. BH-adj. $p < 0.05$ was considered significant (79).

In *study II*, the DEA was performed on the gene counts and protein LFQ values via the R package glmmSeq (105). Linear mixed effects models were built, in which the patient identifiers were included as the random effect. Significance was set at BH-adj. $p < 0.05$ and in case of group comparisons, a minimum of 1.5-fold change (FC) difference was accepted (80).

3.4.5 Pathway analysis

To perform pathway overrepresentation analysis (ORA), either the clusterProfiler (106) and ReactomePA (107) R packages (*study I*), or the fgsea R package was used (*study II*). In *study I*, the default list of human genes was used as background. In *study II*, due to the limited number of quantified proteins and genes, the background gene list varied according to the conducted analysis (either the commonly quantified 162 genes across the tROIs and pROIs were used, or the 2318 proteins and 1811 genes quantified in minimum one pROI and tROI, respectively) (79, 80).

Pre-ranked gene set enrichment analysis (pGSEA) was performed via the clusterProfiler R package. The list of gene sets were downloaded from MSigDB (102) v.7.4 (*study I*) and v.7.5.1 (*study II*), including the Hallmark (99), KEGG (100) and Reactome (101) gene sets which were used in both studies, plus the Gene Ontology (108, 109) biological process and oncogenic curated gene sets which were only used in *study I*. Prior to pGSEA, the proteins/genes were ranked based on a log2(FC) or coefficient value multiplied by the -log10 p-value (all derived from the DEAs described above). In the case of multiple-group comparisons in *study II*, the mean log2(FC)/coefficient was used. The pGSEA p-values were adjusted via the BH method. In *study I*, the pGSEA results were further filtered to obtain the list of subtype-characteristic gene sets. In brief, only pGSEA p-values < 0.01 in all comparisons of the subtype of interest against the other subtypes were accepted, and the normalized enrichment score (NES) had to exhibit an unequivocal sign, either positive or negative. In addition, if a gene set identified as characteristic for a subtype in one dataset (proteomics or transcriptomics) also exhibited a p-value of less than 0.1 in relevant comparisons of the other dataset, and the NES values showed the same direction, the gene set was considered significant in both proteomics and transcriptomics. For visualization purposes, only representative pathways from the list of significant pathways were shown in both studies *I* and *II* (79, 80).

3.4.6 Selection of proteins and genes with stable and variable expression

In *study II*, the coefficient of variation (CV) was calculated for each protein and gene individually within each patient's tumor to calculate the variability of their expression. The ROIs of NATs were excluded from this analysis (and subsequently, proteomic data of Case 2 as it contained only one tumor pROI). Stably and variably expressed proteins/genes were chosen based on whether their CV values were among the bottom and top 20% within at least four cases, respectively (80).

3.4.7 Visualizations

Visualizations were mainly done using the R packages ComplexHeatmap, ggbiplot and ggplot2. Some figures in *study I* were drawn using GraphPad Prism v.8 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com, 2019) (79, 80).

DOI:10.14753/SE.2024.3031

4. RESULTS

4.1 Results from the SCLC study

4.1.1 Cohort overview

We performed a label-free proteomic analysis to characterize the cellular proteome and secretome (CP and CM data, respectively) of 26 cell lines derived from primary or metastatic human SCLCs (Table 1). Using qPCR, these cell lines were categorized into subtypes based on the gene expression patterns of the relevant transcription factors (*ASCL1*, *NEUROD1*, *POU2F3* and *YAP1*). In addition, their culture type was recorded, with 13 growing adherently, 3 semi-adherently, and 10 in suspension (79).

4.1.2 Proteome-level heterogeneity of SCLC cell lines

A total of 10161 proteins were identified and quantified, with 9570 proteins detected in the CP and 6425 in the CM. Among these, 8405 and 5408 proteins were quantified across minimum 80% of the samples in the CP and CM respectively. These filtered data were used for downstream statistical analyses. As expected, the filtered CM data showed a relatively higher number of secreted proteins (n = 514) and proteins detectable in the human blood plasma (n = 3126) compared to the filtered CP data (295 secreted and 3076 plasma proteins) (79).

Cell lines grown adherent on plastic and cell lines grown in suspension showed clearly distinct protein expression profiles. In the CP data, 270 significantly upregulated and 244 downregulated proteins were detected in adherently growing cell lines compared to cell lines grown in suspension. In the CM data, 148 and 244 proteins showed significant upregulation and downregulation in adherently growing cell lines, respectively. Subjecting the differentially expressed proteins separately for CP and CM to pathway ORA showed that KEGG pathways such as the protein processing in endoplasmic reticulum, lysosome and glycosaminoglycan degradation were significantly ($p < 0.05$) enriched both in CP and CM. In addition, other pathways such as the endocytosis and gap junction pathways were found to be overrepresented among the differentially expressed proteins in the CP and CM data, respectively (79).

Multiple observations in the CP data hinted that the cell lines' gene expression-based subtype classification translates into distinct proteomic subtypes. Firstly, the transcription factors showed increased protein-level expression in their respective subtype (Figure 3) (79).

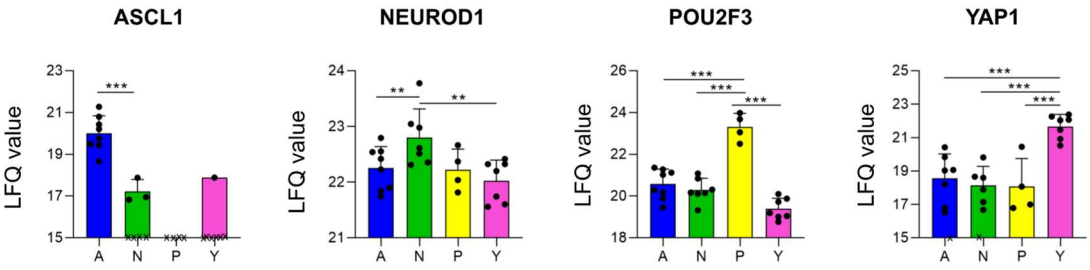

Figure 3. The protein abundance of key transcription factors ASCL1, NEUROD1, POU2F3 and YAP1 derived from the proteomic analysis of small cell lung cancer (SCLC) cell lines. Mean label-free quantification (LFQ) values of proteins ± standard deviation are shown, separately for each subtype (SCLC-A/N/P/Y). Missing LFQ values are indicated by an x. The significance of independent t-tests is indicated above the boxplots (*p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001). Figure in a modified form from (79).

Secondly, an unsupervised consensus clustering of the CP samples based on their most variable proteins grouped the samples into four distinct clusters which corresponded to the qPCR-based subgroups with only one misclassified sample (Figure 4A). The H1882 cell line was classified into the SCLC-A subgroup according to the qPCR results, but into the SCLC-P subset based on the proteomic data. Interestingly, higher *POU2F3* gene expression was detected in this cell line compared to other SCLC-A cell lines. Of note, SCLC-Y cell lines exhibited the most distinct protein expression profile compared to other subtypes. The high concordance between qPCR- and proteome-based subtypes prompted us to use the qPCR-based classification system to group our cell lines into subtypes (79).

Characterizing the cell lines based on their NE and EMT features demonstrated that in most SCLC-A cell lines, NE and epithelial markers were more strongly expressed than non-NE and mesenchymal markers (mean $(M)_{NE}$ score = 0.71, M_{EMT} score = -0.70). SCLC-N cell lines showed NE characteristics with mixed epithelial-mesenchymal characteristics (M_{NE} score = 0.56 and M_{EMT} score = 0.39). Cell lines of the SCLC-P type exhibited lower NE features than SCLC-A and -N but higher than SCLC-Y, together with

DOI:10.14753/SE.2024.3031

a higher expression of epithelial markers (M_{NE} score = -0.05 and M_{EMT} score = -0.61). Lastly, prominent non-NE and mesenchymal features were characteristic for SCLC-Y cell lines (M_{NE} score = -1.35 and M_{EMT} score = 0.75) (Figure 4A). In contrast to the CP data, the CM data were more affected by the culture type of the cell line than by the subtype classification (Figure 4B) (79).

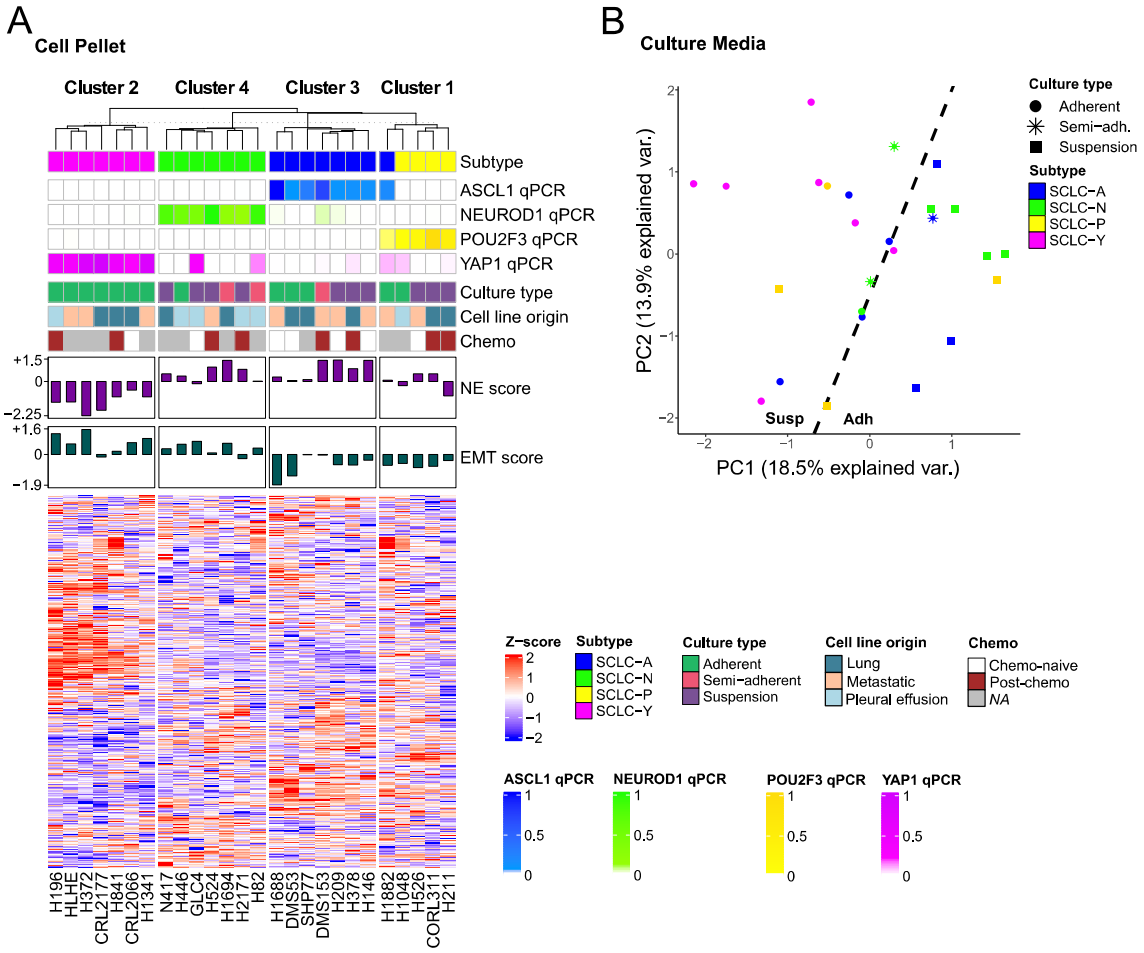

Figure 4. Proteomic analysis of small cell lung cancer (SCLC) cell lines confirms molecular heterogeneity. (A) Heat map of most variable proteins (>1.25 standard deviation, SD) from the cell pellet data. Cell lines are sorted according to their consensus cluster assignments and main sample annotations are shown above the heat map. (B) Principal component analysis plot of the culture media data using only the highly variable (>1.25 SD) proteins, where cell lines are colored according to the subtype and shape indicates culture type. Figures in a modified form from (79).

4.1.3 Proteins with potential diagnostic or therapeutic relevance

To find potential diagnostic (tissue or blood-based) markers for each subtype, proteins showing subtype-specific up- or downregulation (i.e. protein abundance is higher/lower in a given subtype compared to all the three other subtypes) using DEA were sought out. This list was supplemented with four proteins showing on/off characteristics across the subtypes in the CP data, namely achaete-scute homolog 1 (ASCL1; "on" in SCLC-A), regulator of G-protein signaling 22 (RGS22; "on" in SCLC-P), neurexophilin-4 and puratrophin-1 (NXPH4 and PKHG4; "off" in SCLC-Y). In total, 367 and 34 subtype-specific proteins were identified in the CP and CM data, respectively, with 11 proteins displaying subtype-specificity in both datasets. Among these, 33 proteins were assigned to subtype SCLC-A, 54 to SCLC-N, 32 to SCLC-P, and 271 to SCLC-Y. Notably, 22 of these proteins exhibited subtype-specific expression patterns in the CM data and are also detectable in the human blood plasma, thereby representing potential blood-based subtype markers (79).

In addition, six of the subtype-specific proteins were annotated as "druggable" (i.e. are targets of drugs approved by FDA), indicating that these proteins could serve as potential subtype-specific therapeutic targets. The drugs that directly interact with these proteins as part of their mechanism of action were collected (Table 3) (79).

Table 3. List of potentially targetable small cell lung cancer (SCLC) subtype-specific proteins. Abbreviations: CM, culture media; CP, cell pellet; FDA, Food and Drug Administration; RTK, receptor tyrosine kinase; SCLC, small cell lung cancer. Modified table from (79).

Protein name (gene name)	**Dataset & specificity**	**Annotation**	**FDA-approved drugs**
Aromatic-L-amino-acid decarboxylase (*DDC*)	CP, higher expression in SCLC-A	Enzyme that catalyzes dopamine and serotonin synthesis	Benserazide, carbidopa, methyldopa
Ephrin type-A receptor 2 (*EPHA2*)	CP, higher expression in SCLC-Y	RTK, involved in contact-dependent bidirectional signaling with neighboring cells	Dasatinib, regorafenib
Histone deacetylase 1 (*HDAC1*)	CP, higher expression in SCLC-A/N/P	Histone deacetylase with regulatory function in transcriptional processes	Romidepsin, vorinostat
Integrin alpha-V (*ITGAV*)	CP, higher expression in SCLC-Y	Integrin, receptor for a wide array of proteins. Cluster of differentiation marker	Antithymocyte immunoglobulin, levothyroxine
Integrin beta-1 (*ITGB1*)	CP, higher expression in SCLC-Y	Integrin, receptor for a wide array of proteins. Cluster of differentiation marker	Antithymocyte immunoglobulin
Mast/stem cell growth factor receptor kit (*KIT*)	CP and CM, higher expression in SCLC-P	RTK, acts as cell-surface receptor for the cytokine KIT Ligand. Cluster of differentiation marker	Ancestim, imatinib, lenvatinib, pazopanib, regorafenib, ripretinib, sorafenib, sunitinib, tivozanib

Several SCLC cell lines have previously been tested against seven of these drugs (CancerRxGene database, (91)). Contrasting the drug sensitivity data with our proteomic data showed that lower ephrin type-A receptor 2 (*EPHA2*), mast/stem cell growth factor receptor kit (*KIT*) and histone deacetylase 1 (*HDAC1*) protein abundance were indicative of increased resistance (i.e. generally higher logarithmic half maximal inhibitory concentration (IC50) values) to the drugs (dasatinib, pazopanib, and vorinostat respectively) targeting these proteins (Figure 5) (79).

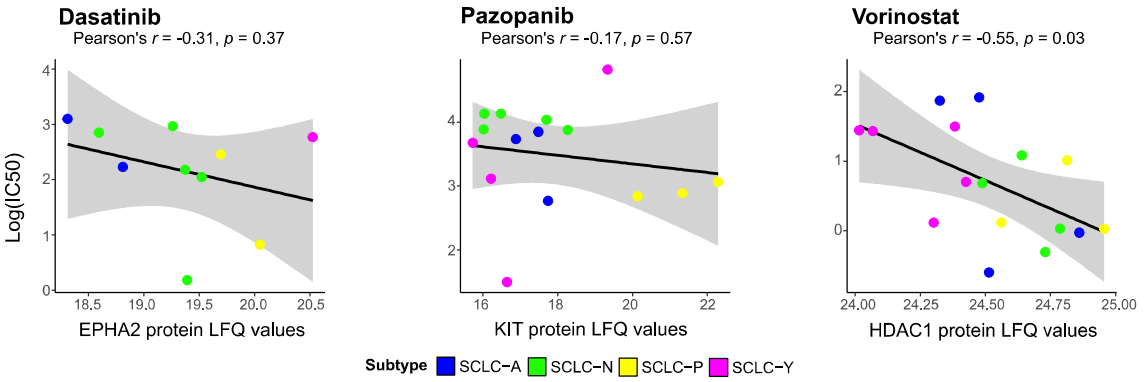

Figure 5. Relationships between protein expression and drug sensitivity of small cell lung cancer (SCLC) cell lines. Scatter plots showing the logarithmic half maximal inhibitory concentration (IC50) values of the cell lines for drugs selected from the Genomics of Drug Sensitivity in Cancer 1 (GDSC1) database (from left to right: dasatinib, pazopanib, and vorinostat), as a function of the measured protein label-free quantification (LFQ) values (EPHA2, KIT, HDAC1). Pearson correlation analysis results are indicated above the plots, and dots are colored according to the cell lines' subtype assignment. Figure in a modified form from (79).

4.1.4 The multi-omic portraits of SCLC subtypes

To detect the unique pathway-level patterns for each subtype, pGSEAs were performed for each pairwise subtype comparison both in our CP data and in the RNA-Seq data of the 50 SCLC cell lines from CCLE (94). The subtype-specificity of the pathways was additionally validated using the SCLC tissue RNA-Seq data published by George et al. (18). In this validation process, ssGSEA scores for the selected pathways were calculated for each tumor individually. Subsequently, these scores were averaged per subtype to assess the pathway activity trends across the subtypes (79).

In the SCLC-A subtype, upregulation of members of the neural precursor cell proliferation was supported by both the proteomic and transcriptomic cell line dataset (Figure 6A). In addition, oxidative phosphorylation (OXPHOS) and respiratory chain elements displayed concordant overexpression in SCLC-A according to proteomics, while the members of the subpallium development gene set were only significantly upregulated according to transcriptomics. Of note, the higher expression of OXPHOS elements in this subtype was also supported by the SCLC tissue transcriptomics data (79).

Regarding SCLC-N, subtype-specific downregulation of epidermis development process elements was observable in both cell line proteomics and transcriptomics data (Figure 6B). The proteome of SCLC-N could be further characterized by the

downregulation of immune response, cytokine signaling, cytoskeleton organization and cell adhesion-related proteins, as well as by the upregulation of DNA replication and transcription-associated proteins. While not confirmed with cell line transcriptomics, several of the aforementioned pathways showed the subtype-specific trend in the tissue transcriptomics data (79).

Considering SCLC-P specific pathways, elements of the neurotrophin signaling pathway and the lamellipodium organization gene set were detected as concordantly overexpressed in SCLC-P, supported by both cell line dataset but not confirmed by tissue transcriptomics (Figure 6C) (79).

Lastly, a substantial number of processes showed SCLC-Y-specific patterns (Figure 6D). In particular, members of extracellular matrix (ECM) organization, cytokine-mediated signaling, interleukin signaling, inflammatory response, EMT, cell-substrate adhesion, response to growth factors, and MAPK cascade were overexpressed in SCLC-Y compared to other subtypes, both in the cell line proteomic and transcriptomic data. With the exception of cytokine-mediated signaling, all subtype-specific trends were validated by tissue transcriptomics. At the gene expression level, upregulation of apoptotic pathway and the JAK-STAT signaling members was observed. On the other hand, proteomics revealed the higher expression of proteins associated with the signaling by Rho-GTPases, as well as with the activation of transmembrane transporter disorder-related processes. Moreover, protein-level data showed the downregulation of protein acetylation, DNA repair and chromatin modification elements in this subtype. The latter two gene sets showed the same subtype-specific trend in tissue transcriptomics as well (79).

DOI:10.14753/SE.2024.3031

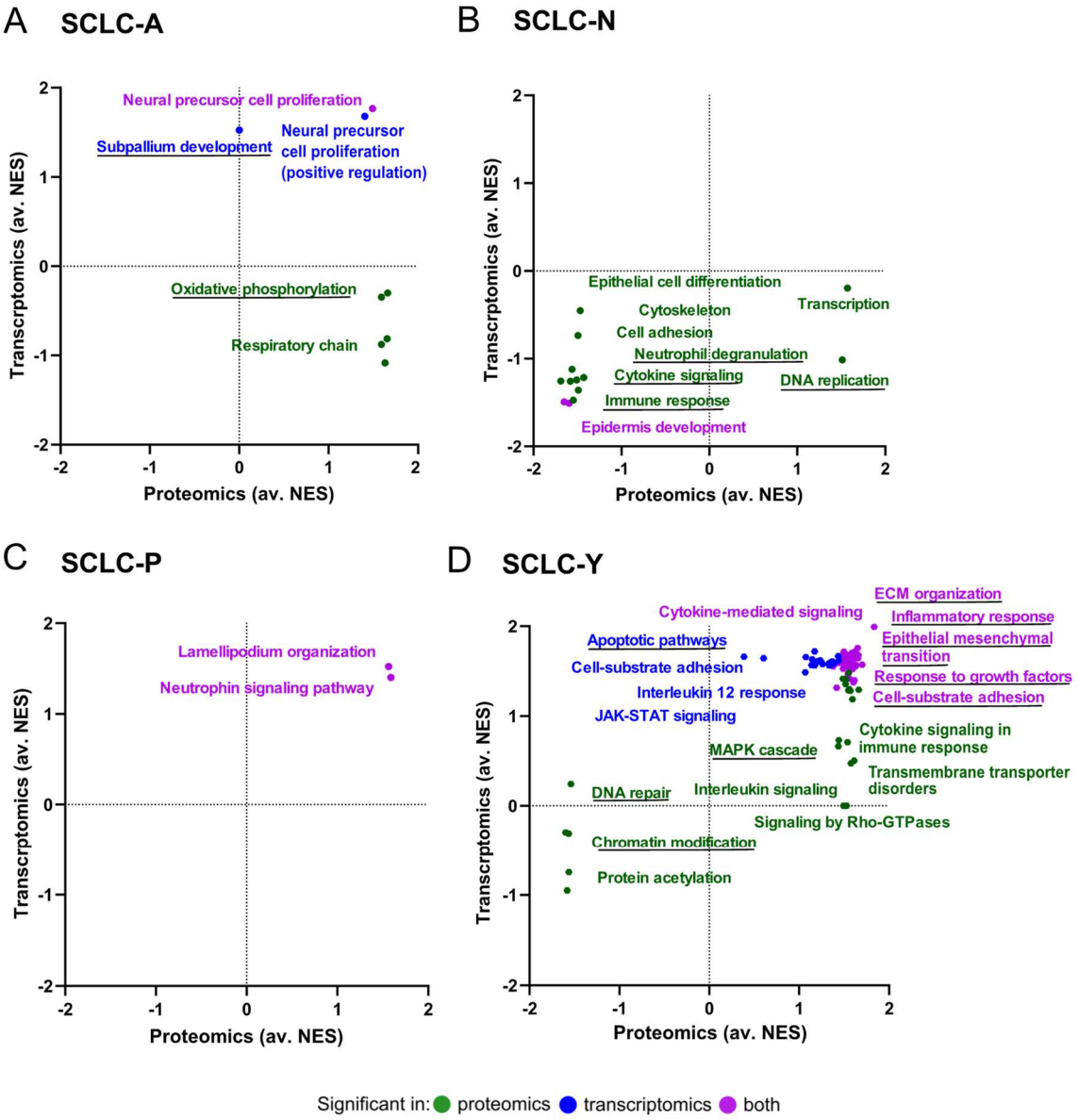

Figure 6. Subtype-specific biological processes detected in small cell lung cancer (SCLC) subtypes SCLC-A/N/P/Y. (A) SCLC-A-specific processes. (B) SCLC-N-specific processes. (C) SCLC-P-specific processes. (D) SCLC-Y-specific processes. The subtype-specificity of underlined pathways was supported by the SCLC tissue transcriptomic data published by George et al. (18). Abbreviations: avr., average; NES, normalized enrichment score. Figures in a modified form from (79).

DOI:10.14753/SE.2024.3031

4.2 Results from the *ALK*-rearranged LADC study

4.2.1 Cohort overview

Seven primary LADC tumors with confirmed *ALK* rearrangement were investigated in this study (Table 2). The tumor samples were collected before any therapeutic interventions, and all patients underwent treatment with ALK-inhibitors following the confirmation of *ALK* positivity. Two out of seven patients passed away due to lung cancer until the last follow-up (January/February 2022). Selected tumor and NAT regions on the FFPE tissues were subjected to transcriptomic and proteomic analyses (Figure 2). The tROIs (n = 84) were chosen by taking into account their morphologic setting and varying levels of lymphocytic infiltration. In contrast, the pROIs (n = 23) covered greater areas of the whole tumor slide (80).

4.2.2 Multi-omic heterogeneity of *ALK*-rearranged LADCs

Through NanoString GeoMx Digital Spatial Profiling, a total of 1811 genes were quantified across all tROIs. Label-free proteomics resulted in the quantification of 2318 proteins in at least one pROI, out of which 1154 proteins were quantified across minimum 80% of the pROIs and were kept for statistical analyses (80).

The mean Pearson correlation coefficient (r) between proteomics and transcriptomics was 0.43±0.33 at the gene level (n = 162 genes), with 69 and 1 genes showing significant positive and negative correlation respectively. At the singscore level (n = 431 gene sets), the mean Pearson's r between pROI and tROI data was 0.24±0.34, and 63 and 6 gene sets showed a significant positive and negative correlation, respectively. Positive correlation between proteomics and transcriptomics was detected for pathways including metabolism (e.g., glycolysis, RNA, tyrosine and phenylalanine metabolism), immune-system-related processes (such as adaptive immune system, allograft rejection), signal transduction and gene regulation (mTORC1 signaling, MYC targets, p53 pathway, NCAM signaling), cell cycle and DNA replication. The only significantly negatively correlated gene across the two molecular layers was *RPLP0*. In addition, the negatively correlating processes (assessed based on singscores) included MAP2K and MAPK activation, oncogenic MAPK signaling, and developmental biology (80).

DOI:10.14753/SE.2024.3031

In addition, unsupervised consensus clustering of the pROIs and tROIs (Figure 7A-B) revealed that clustering was largely driven by interpatient differences with the exception of NAT regions, which were clustering together regardless of patient (80).

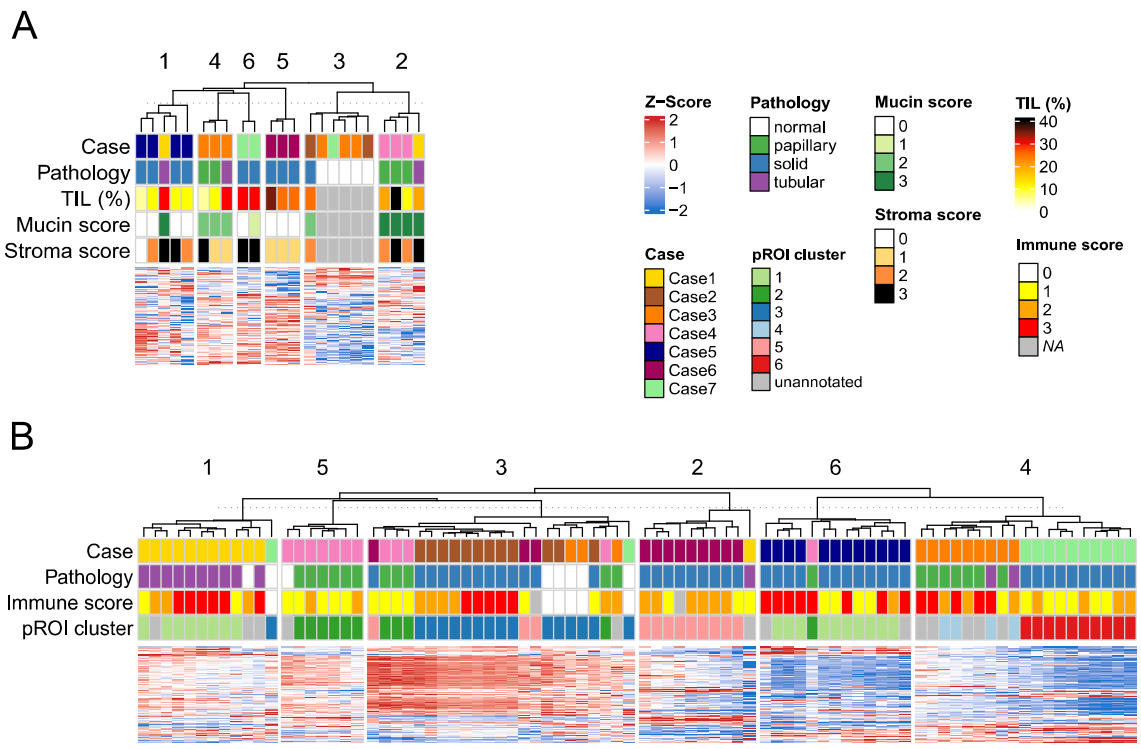

Figure 7. The observed sample heterogeneity of the seven *ALK*-rearranged lung adenocarcinomas (LADCs). (A) Heat map showing consensus clustering results for the proteomic regions of interest (pROIs) based on the proteins' Z-scored label-free quantification (LFQ) values. (B) Heat map showing consensus clustering results for the transcriptomic regions of interest (tROIs) based on the Z-scored gene counts. Figures in a modified form from (80).

4.2.3 Molecular changes associated with histopathological characteristics

Comparing tumor and NAT regions with DEA, a total of 310 proteins and 47 genes were significantly upregulated in tumor pROIs and tROIs respectively (from which 10 were shared between pROI and tROI results), while 136 proteins and 38 genes were upregulated in NAT pROI and tROI regions respectively (from which 3 were shared between the two datasets). Notably, the *C3*, *C4BPA*, *CLU*, *COL1A1* and *LAMB3* genes showed upregulation in tumors (tROI data), but their protein products showed downregulation in tumors (pROI data). Interestingly, when examining the list of differentially expressed genes and proteins in the LADC dataset from CPTAC (7), only the *ALK*-rearranged tumor and NAT subset of the CPTAC cohort confirmed the trends

observed in our data for glutathione peroxidase 1 (*GPX1*, upregulated in tumors) and complement component 4 binding protein alpha (*C4BPA*; conflicting differential expression patterns across proteomics and transcriptomics) (Figure 8A) (80).

At the pathway level, supported by both molecular layers, tumor regions displayed an increased expression of members of glycolysis, mTORC1 signaling, unfolded protein response and infection pathways compared to NAT regions, as uncovered by pGSEA (Figure 8B). The proteomic data highlighted additionally the upregulation of members of translation, RNA, glucose, and amino acid metabolisms in tumors, as well as the downregulation of proteins involved in apical junction, ECM organization, receptor tyrosine kinase (RTK) signaling such as MAPK signaling and Toll-like receptors. Of note, the members of the complement and coagulation cascade were upregulated in tumors according to the tROI data but downregulated according to the pROI data (80).

Regarding immune infiltration patterns, the sulfotransferase family 1A member 1 (*SULT1A1*) protein showed significant upregulation with increasing tumor infiltrating lymphocyte (TIL) amounts in the pROI data, while the tROI data showed the upregulation of the B alpha chain of HLA class I histocompatibility antigen (*HLA-B*) and the downregulation of prostaglandin-endoperoxide synthase 2 (*PTGS2*) with higher immune score. Only the tROIs showed significant pathway-level differences with increasing immune score according to pGSEA. Consistent with pathological evaluation, the list of significantly activated pathways with higher immune score in the tROI data were mainly immune system related. Notable examples include antigen processing and presentation, interferon signaling and neutrophil degranulation (80).

DOI:10.14753/SE.2024.3031

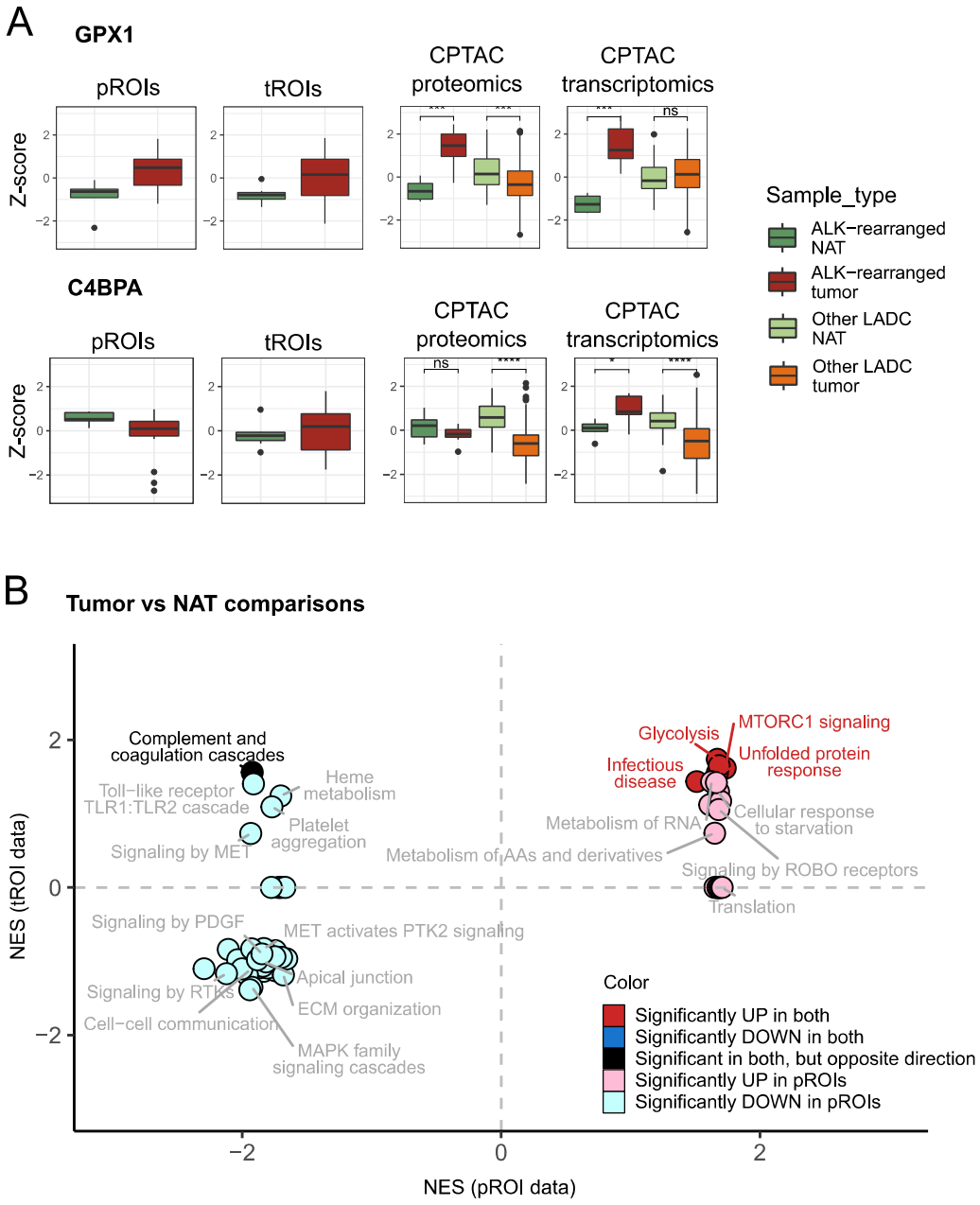

Figure 8. Proteomic and transcriptomic differences between tumor and normal adjacent tissue (NATs) regions in *ALK*-rearranged lung adenocarcinomas (LADCs). (A) Tumor vs NAT expression differences for *GPX1* and *C4BPA* in our data, validated by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) LADC study (7). Symbols indicate t-test significance (ns: not significant, *p-value ≤ 0.05; **p-value ≤ 0.01; ***p-value ≤ 0.001). (B) Normalized enrichment scores (NESs) derived from pre-ranked gene set enrichment analysis for tumor vs NAT comparisons. NES from the analysis of proteomic regions of interest (pROIs) and transcriptomic regions of interest (tROIs) are shown on the x- and y-axis respectively. Figures in a modified form from (80).

In the pROI data, we additionally investigated mucin and stroma score-related proteomic differences. In total, 4 and 13 proteins were upregulated and downregulated with increasing mucin score, while 1 and 3 proteins were upregulated and downregulated with increasing stroma score, respectively. At the pathway level, increasing mucin and stroma scores could be associated with the concordant upregulation of EMT-related proteins and ECM organization. Parallelly, members of RNA metabolism or signaling by ROBO receptors were downregulated with higher mucin and stroma scores (Figure 9) (80).

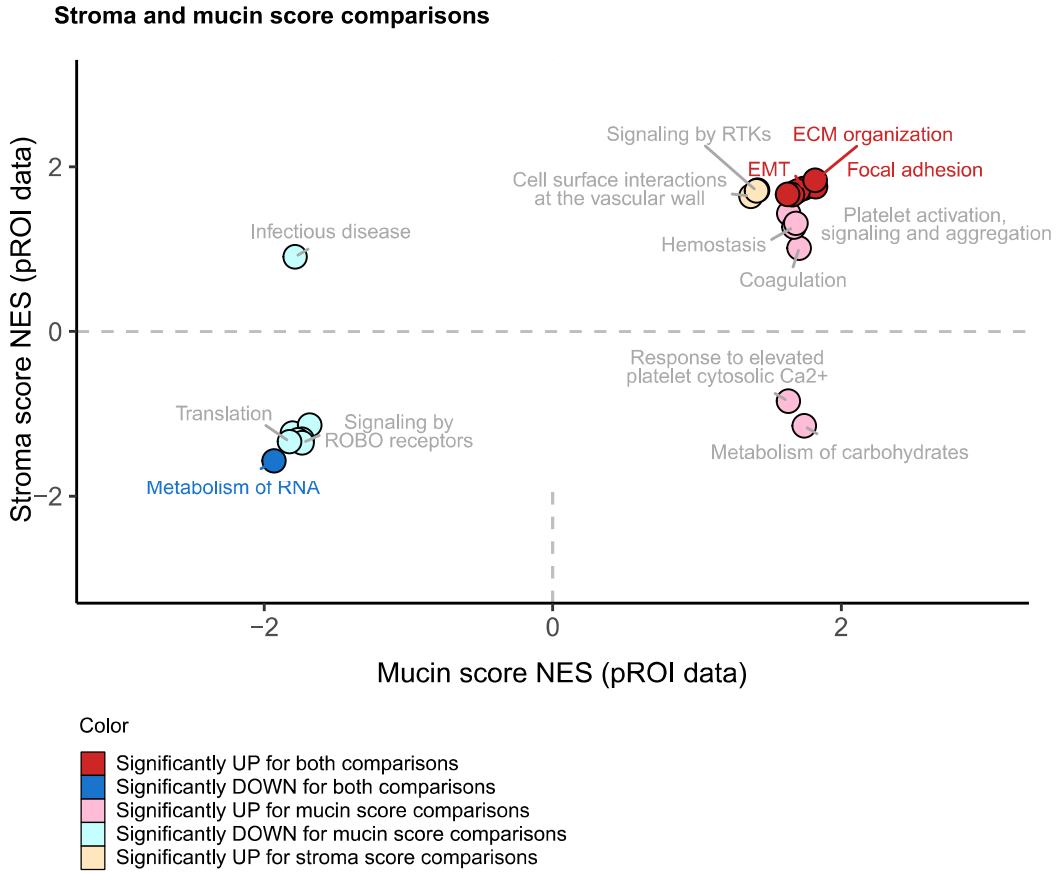

Figure 9. Proteomic differences linked to increasing mucin and stroma scores in *ALK*-rearranged lung adenocarcinoma (LADC) tumor regions. Normalized enrichment scores (NESs) derived from pre-ranked gene set enrichment analysis results for the mucin high vs low (x-axis) and stroma high vs low (y-axis) comparisons in the proteomic regions of interest (pROI) data are shown. Figure in a modified form from (80).

4.2.4 Intratumoral homogeneity and heterogeneity

To uncover the drivers of intratumoral molecular homogeneity and heterogeneity, the overall expression variability for each protein and gene within each case was assessed by computing their CV. Notably, three heat shock proteins (encoded by genes *HSPA1A*, *HSPB1* and *HSP90B1*) showed high stability at the protein expression level within minimum four tumors, but high variability at the gene expression level. In parallel, both proteins and genes with a stable expression within minimum four tumors showed an enrichment (ORA $p < 0.05$) for PI3K-AKT-mTOR signaling, oncogenic MAPK signaling, and "signaling by moderate kinase activity BRAF mutants" pathways (Table 4) (80).

The pathway analysis for the proteins and genes showing high variability (n = 101 and 232, respectively) within minimum four tumors resulted in the enrichment of gene sets such as the ECM-associated pathways, EMT, complement and coagulation cascades, MET-activated PTK2 signaling, angiogenesis and glycolysis (ORA $p < 0.05$), supported by both molecular layers (Table 4). Among the proteins and genes showing high variability across all examined cases (six cases for pROIs and seven for tROIs), fibronectin 1 (*FN1*) stood out for being highly variable at both molecular layers. Interestingly, this protein is connected to several pathways enriched for heterogeneously expressed genes and proteins, indicating its potential significance in the context of the observed molecular heterogeneity (80).

Table 4. Pathways enriched for proteins and genes that are stably or variably expressed across tumor ROIs. Only pathways are mentioned which were commonly enriched both in the pROI and tROI data with $p < 0.05$. Abbreviations: adj, adjusted; ECM, extracellular matrix; nr, number; pROI, proteomic region of interest; tROI, transcriptomic region of interest. Data from (80).

Enriched pathways	**pROI data**		**tROI data**	
	Nr. of proteins in pathway	**Adj. p for pathway enrichment**	**Nr. of genes in pathway**	**Adj. p for pathway enrichment**
for stably expressed proteins/genes				
PI3K-AKT-mTOR signaling	7	3.9×10^{-4}	15	1.1×10^{-1}
Oncogenic MAPK signaling	5	9.1×10^{-2}	10	1.1×10^{-2}
Signaling by moderate kinase activity BRAF mutants	4	1.3×10^{-1}	8	1.7×10^{-2}
for variably expressed proteins/genes				
Angiogenesis	4	8.6×10^{-2}	9	1.6×10^{-2}
EMT	15	1.4×10^{-3}	34	1.3×10^{-8}
Complement and coagulation cascades	7	6.0×10^{-2}	19	4.3×10^{-6}
ECM-receptor interaction	9	2.8×10^{-2}	17	5.6×10^{-3}
Degradation of the ECM	12	3.7×10^{-3}	16	8.9×10^{-4}
Diseases of glycosylation	6	1.1×10^{-1}	5	1.7×10^{-1}
ECM proteoglycans	12	5.9×10^{-5}	13	2.2×10^{-2}
ECM organization	24	3.1×10^{-7}	29	3.0×10^{-3}
Integrin cell surface-interactions	9	2.9×10^{-2}	17	6.6×10^{-3}
MET activates PTK2 signaling	6	2.8×10^{-2}	11	3.0×10^{-3}
MET promotes cell motility	6	6.0×10^{-2}	11	5.6×10^{-3}
Non-integrin membrane-ECM interactions	6	1.9×10^{-1}	11	7.3×10^{-2}
Regulation of IGF transport and uptake by IGF-binding proteins	9	2.9×10^{-2}	11	1.1×10^{-2}
Response to elevated platelet cytosolic Ca^{2+}	8	2.4×10^{-1}	13	4.2×10^{-2}
Signaling by MET	6	8.6×10^{-2}	11	1.6×10^{-1}

DOI:10.14753/SE.2024.3031

5. DISCUSSION

5.1 Study of SCLC subtypes

Molecular heterogeneity in SCLC is currently under intense investigation given the subtypes' prognostic and therapeutic relevance (110). Next to the SCLC cell lines with classical NE phenotype and suspension growth type, a "variant" form (NE-low) of SCLC, mainly forming adherent cell cultures, as well as tumors lacking NE differentiation have been reported (16, 111, 112). In our study, half of the SCLC cell lines grew adherently, and in line with previous findings, this growth pattern was associated with non-NE characteristics. We linked distinct protein profiles to adherently versus in-suspension-growing cell lines, such as dysregulation of glycosaminoglycan degradation, endocytosis, and gap junction pathways (79).

According to a classification system proposed by Rudin et al. (16), our cell lines were classified into SCLC-A, -N, -P and -Y subtypes based on the mRNA expression patterns of *ASCL1*, *NEUROD1*, *POU2F3* and *YAP1* (n = 8, 7, 4 and 7, respectively). At protein level, these transcription factors exhibited higher abundances in their respective subtypes as well. Based on the protein expression of NE and non-NE markers, we confirmed that SCLC-A and -N cell lines exhibit NE features, while SCLC-P and -Y groups show rather non-NE features (79). Recent IHC-based studies on human tumor tissues corroborated the overexpression of NE markers in SCLC-A and -N, while SCLC-P tumors did not display notable NE differentiation at the protein level (21, 113). Regarding epithelial and mesenchymal features, a previous SCLC cell line study already outlined that SCLC-P and -Y exhibit epithelial and mesenchymal attributes, respectively, while SCLC-A and -N display mixed characteristics (114), which was in line with our findings (79).

Importantly, according to unsupervised cluster analysis, the SCLC cell lines could be categorized into four subgroups based on their CP proteomic profiles, which was in alignment with their predefined mRNA-based subtypes (except for one misclassified cell line). The overall high variability of protein expressions in the CM data compared to the CP data (mean SD of 1.20 and 0.85 in the CM and CP data respectively), may have affected the detection of subtype-specific signatures in the CM proteome. Another

contributing factor to proteomic differences across cell lines was the culture type, impacting both the CP and CM (79).

We report 367 and 34 proteins in the CP and CM showing subtype-specific expression patterns, among which six proteins are also FDA-approved drug targets. Specifically, DDC was found to be overexpressed in SCLC-A, while EPHA2, ITAV, and ITB1 (corresponding to genes *EPHA2*, *ITGAV*, and *ITGB1*) were overexpressed in SCLC-Y. HDAC1 was downregulated in SCLC-Y, and KIT was upregulated in SCLC-P (79). High levels of *ASCL1* correlated with stronger expression of *DDC* (115), and *EPHA2*, a non-NE marker (17), was previously reported to be upregulated in SCLC-Y (114). The overexpression of integrins in SCLC-Y has been linked to chemotherapy resistance through the suppression of chemotherapy-induced apoptosis (116). In agreement with this, we previously reported a positive correlation between YAP1 protein abundance and resistance to chemotherapeutic agents in SCLC cell lines (110). The downregulation of HDAC1 in SCLC-Y in our study aligns with the reported HDAC inhibitor resistance in SCLC-Y (117). KIT protein is a known marker for SCLC-P (118). Additionally, our hypothesis that overexpression of proteins in one or multiple subtypes may indicate sensitivity of those subtypes to drugs targeting these proteins was supported by CancerRxGene data (91) and our follow-up study (119). Specifically in the present study, we confirmed that KIT-targeting drugs, such as pazopanib, are potentially suitable for targeting *POU2F3*-driven SCLCs, while *YAP1*-driven tumors are more resistant to vorinostat, a drug that targets HDAC1 protein (79).

To delineate the distinctive characteristics of SCLC subtypes, we also conducted a multi-omic pathway-level analysis, integrating expression differences from both quantified proteins in our dataset and transcripts from CCLE transcriptomic data (94), supplemented by observations from an SCLC tissue transcriptome data (18). This comprehensive approach revealed a list of potential subtype-specific therapeutic vulnerabilities (79).

Given *ASCL1*'s established regulatory role in neural differentiation (120), neural precursor cell proliferation and subpallium development-related proteins showed a concordant upregulation in the SCLC-A subtype. Our proteomic data strongly indicated that the activation of OXPHOS and respiratory chain elements is highly specific to SCLC-A (79). Cell lines lacking *MYC* expression, which is characteristic of SCLC-A (16, 18),

tend to rely more on oxidative metabolism, suggesting potential susceptibility of SCLC-A tumors to OXPHOS inhibitors (79).

In our study, SCLC-N cell lines, primarily forming suspension cultures, showed a consistent downregulation of cell adhesion pathways. Proteomic data also revealed increased activity in DNA replication and transcription, coupled with a decrease in cytokine-mediated signaling in this subtype (79). Notably, a similar trend was observed in a study comparing gene expression between SCLC and normal tissue (121).

Epithelial-like SCLC-P cell lines displayed an upregulation in the lamellipodium organization pathway, a critical step in EMT associated with enhanced cell motility and invasive capacities (122). In addition, the observed upregulation in neurotrophin signaling suggests the potential effectiveness of PARP inhibitors as therapeutic agents for SCLC-P, as previously proposed (22), given that prolonged PARP activation can contribute to neurotrophin-induced neuronal death (123). Additionally, our data implies that direct targeting of neurotrophin signaling could be a suitable treatment option in *POU2F3*-driven SCLC (79).

SCLC-Y cell lines formed a distinct subgroup in our samples. Proteomics revealed a unique downregulation of protein acetylation, chromatin modification, and DNA double-strand break repair pathways in this subtype, in parallel with the overexpression of the MAPK cascade and Rho-GTPase signaling members (the latter supported by the known role of Rho in YAP/TAZ activity (124)) (79). In a previous study, SCLC-A was found to be selectively sensitive to MAPK activation compared to SCLC-N and -P, but no cell lines from the SCLC-Y subtype were tested by the authors (125). Due to the adherent nature of *YAP1*-driven cell lines, we observed an upregulation of focal adhesion, ECM organization, and cell-substrate protein pathways along with peroxisome and endocytosis-related proteins in the SCLC-Y subgroup (79). Tlemsani et al. highlighted that SCLC-Y cell lines exhibit high presenting and native immune predisposition, and have the highest antigen-presenting machinery scores, thus we can anticipate sensitivity to immune-checkpoint inhibitors (114). Consistent with this, we identified a distinctive upregulation of cytokine-mediated signaling and inflammatory response in this subtype (79).

There are controversies surrounding the existence of the SCLC-Y subtype, as comprehensive immunohistochemical and histopathological analyses of SCLC subtypes

in patient samples have not identified a distinct *YAP1*-driven subtype (21). Conversely, our preclinical proteomic study clearly delineates a distinct SCLC-Y subtype among the examined cell lines (79).

Cell lines enable the examination of pure populations of homogeneous tumor cells without the presence of admixed stromal or inflammatory cells, which is particularly valuable in drug sensitivity assays and subtyping studies, such as the current one (126, 127). However, some study limitations need to be mentioned. The number of examined cell lines was low (n = 26), and the proteomic profiles of the adherent culture types and SCLC-Y could not be investigated independently because all *YAP1*-driven cell lines grew adherently. The treatment-related data about the original tumors was not retrievable for some SCLC cell lines, partly because the majority of the examined cell lines were established in the past century. However, it is noteworthy that neither the NE/mesenchymal characteristics nor the protein expression profile differed significantly based on the presence or absence of chemotherapy (data not shown). It is also notable that unique up- or downregulation of pathways in an SCLC subtype may not necessarily translate to dependency or independency from those pathways due to the interconnected nature of biological processes and numerous regulatory factors, including feedback and feed-forward loops which interfere with such processes. In conclusion, our results should primarily be viewed as hypothesis-generating for future studies. Subsequent proteomic analyses of larger SCLC cohorts, preferably incorporating fresh tissue samples from patients with homogeneous treatment histories, are needed to validate our findings (79).

5.2 Study of *ALK*-rearranged LADCs

Spatial molecular profiling of tumors is increasingly popular and enhances our understanding of cancer, including lung cancer, by providing the crucial spatial context (40-43). We explored the proteome and transcriptome-level drivers of inter- and intratumoral heterogeneity in seven LADCs with confirmed *ALK* rearrangements, which were collected prior to treatment with ALK inhibitors (80).

From the FFPE slides, larger proteomic and smaller transcriptomic regions including NAT ROIs and tumor ROIs with various histopathological features (morphology, immune infiltration, mucin and stroma scores) were selected for molecular profiling. Morphological and other histological features are often combined within a

LADC tumor and carry prognostic value (30, 128). In this study, only one tumor displayed multiple morphologies. Notably, a multitude of regions (10 out of 23 pROIs and 44 out of 84 tROIs) showed solid patterns, which was accompanied by lower mucin scores compared to papillary and tubular morphologies (80).

We achieved the confident quantification of 1154 protein groups and 1811 genes across the pROIs and tROIs, and only 162 genes were commonly present in both data sets (80). A lower positive correlation between the two molecular layers (median Pearson's r = 0.43 and 0.24 at gene and pathway levels, respectively) was observed, which was in line with previous studies reporting moderate (7, 8, 10) to low (6, 129-131) positive correlations between proteomics and transcriptomics. In our study, positive correlations were present in pathways such as amino acid metabolism, glycolysis, p53 pathway, adaptive immune system, and DNA replication (80). This was consistent with a previous finding by Chen et al. (6). The significant negative correlation we detected for the ribosomal protein RPLP0 was supported by the previous observations that ribosomal functions tend to be lowly correlated between the two molecular layers (10, 131, 132). The poor correlation for some RNA and protein abundances, e.g., for members of developmental biology and MAPK signaling pathway (80), may result from post-transcriptional or -translational regulation not captured by our study (133-135).

Unsupervised clustering of pROIs and tROIs revealed that intertumoral heterogeneity is more pronounced than intratumoral heterogeneity, except for NAT regions, which were distinct based on their molecular expression profiles. Comparing tumor and NAT regions resulted in numerous differentially expressed proteins and genes, with some overlap across the molecular layers. Notably, three complement cascade proteins (corresponding to genes *C3*, *C4BPA*, and *CLU*) and two proteins secreted to the ECM (encoded by genes *COL1A1* and *LAMB3*) showed opposite tendencies between pROIs and tROIs (downregulated in tumors at pROI level but upregulated in tumors at the tROI level compared to NATs). Interestingly, we observed *GPX1* showing higher gene and protein expressions in ALK-rearranged LADCs compared to NATs (80), which was not detectable in *ALK*-negative LADCs of the CPTAC cohort (7). Similarly, another study not focusing on the *ALK*-driven subgroup identified *GPX1* as downregulated in lung tumors compared to NATs (136). The role of this gene in lung cancer is not yet elucidated

(137), however, one study indicated that *GPX1* may induce cisplatin-based chemoresistance in NSCLC (138).

Pathway-level differences between tumors and NATs reflected known cancer hallmarks, such as impaired glycolysis (139), unfolded protein response (140), translation (141), ECM organization (142) pathways or signaling by RTKs (143). The last three hallmarks were only significant in the proteomic data, which underscores the significance of proteomics in spatial omic studies (80).

The investigation on the role of tumor-infiltrating immune cells and potential immunotherapy strategies in NSCLC is ongoing (40-43). It is noteworthy to explore these aspects in the *ALK*-rearranged subtype as well, where immune-based therapies have shown limited efficacy (144). Investigating immune-infiltration-related patterns in our pROI data showed SULT1A1 protein being upregulated with increasing TIL %. Regarding tROI expression data, *HLA-B* and *PTGS2* genes exhibited up- and downregulation with increasing immune score respectively. These 3 proteins/genes represent potential markers for variability in immune infiltration. Additionally, multiple immune-related processes, such as neutrophil degranulation, or antigen processing and presentation, , were upregulated with increasing immune scores in the tROIs, confirming agreement between the gene expression signatures and pathological evaluation. In contrast, no pathways were significant when comparing pROI regions with varying TIL%, possibly due to more heterogeneous immune infiltration patterns within the pROIs (80).

Our previous publication on the hereby presented proteomic data showed the mucin content's significant impact on proteomic and glycosaminoglycan profiles while the effect of stromal content was less prominent (83). This finding was reaffirmed in our current analysis, where we observed a greater number of differentially expressed proteins associated with mucin score compared to stroma score (17 vs 4), with no overlap in the lists of significant proteins. However, despite these differences, we observed that alterations in mucin and stroma scores across the pROIs affected similar pathways. For example, upregulation of ECM components was detected both with higher mucin and stroma scores, aligning molecular findings with histologically visible phenotypes (80).

Focusing on intratumoral molecular homogeneity, our data showed that members of the oncogenic MAPK and PI3K-AKT-mTOR signaling generally show stable

expression across tumor regions at both proteome and transcriptome level (80). This is consistent with the observation that aberrant *ALK* activity leads to the activation of both pathways (145). Investigation of pathways that can contribute to intratumoral heterogeneity pointed to EMT and ECM organization (80). EMT is one of the cancer hallmarks (146) which may mediate ALKi resistance (147). Within the EMT process, ECM remodeling plays an essential role and promotes cancer metastasis (148). Notably, three heat shock proteins displayed high expression stability at the protein level but high expression variability at the gene level (80). Heat shock proteins are produced when cells are under stressful conditions, thus the observed disagreement may stem from post-transcriptional regulation enabling cells to adapt to stress in a timely manner (149). Interestingly, a well-known EMT marker, FN1, showed high intratumoral variability across ROIs at both protein and gene level (80). FN1 can have both tumor-promoting and -suppressive characteristics (150), and some studies link lower FN1 expression to a more favorable outcome (151, 152), while others noted the downregulation of FN1 with LADC progression (153, 154) or no relationship with survival (155, 156). The heterogeneous FN1 expression within tumor tissue, as demonstrated in this study, may contribute to these controversies in the literature. Moreover, there is a lack of investigations into the specific role of FN1 in *ALK*-rearranged LADCs (80).

Limitations of this study need to be noted. While we strengthened the reliability of our findings by validating multiple results with data from a larger, previously published LADC cohort (CPTAC) (7), the observations made on this small cohort with heterogeneous clinical data may not be generalizable to all *ALK*-driven LADCs. Evaluating individual molecular characteristics of histopathological features (such as mucin and stroma content) is challenging due to their interconnected nature. Intratumoral molecular variability may be influenced by stochastic factors at the cellular level and phenotypical differences, as well as varying extent of tumor purity. Both the spatial gene expression profiling and the shotgun proteomic approach have identification and quantification biases (bias for known cancer-related genes and high-abundant proteins, respectively) (80). These are unavoidable technical and biological limitations which often lead to milder correlations in the measured protein and transcript abundances (133-135). In addition to this, the size-differences in the examined pROIs and tROIs in our study may also impact the agreement between proteome- and transcriptome-level results (80).

The pROI sizes are inherent to the technique used in this work (on-tissue digestion with repeated pipetting), providing the required efficiency and reproducibility (85). Overall, a larger cohort of *ALK*-rearranged LADCs will be necessary to investigate in the future to corroborate the findings presented in this study (80).

DOI:10.14753/SE.2024.3031

6. 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.

DOI:10.14753/SE.2024.3031

7. SUMMARY

Advances in omics profiling have enhanced our understanding of lung cancer, revealing new molecular subgroups and new avenues in targeted therapies. Proteomics, in particular, can bring new insights into cancer by offering a closer representation of the phenotype. In our work, we aimed to perform a proteomic analysis complemented with gene expression data on relatively rare lung cancer types to gather insights on their intra- and intertumoral heterogeneity.

In the small cell lung cancer (SCLC) study, we focused on characterizing newly defined SCLC subtypes at the protein level. We examined pellets and cell media of 26 human SCLC cell lines, supplemented with public transcriptomic data of SCLC cell lines and tissues. We confirmed the presence of molecular subtypes at protein level which were previously defined based on the gene expression of key transcription regulators: *ASCL1*, *NEUROD1*, *POU2F3*, and *YAP1*. We identified proteins with potential subtype-specific expression, including known druggable proteins and potential blood-based markers, and pathways that may be uniquely activated or suppressed in a certain subtype compared to the other three subtypes, representing further potential therapeutic targets. Importantly, while tissue-based immunohistochemistry studies failed to confirm a *YAP1*-driven subtype, we found it to be the most distinct SCLC subgroup.

In the study of *ALK*-rearranged lung adenocarcinoma (LADC), we performed, for the first time, spatial transcriptomic (NanoString GeoMx, 12 regions per case) and proteomic profiling (2-6 regions per case) on seven treatment-naïve formalin-fixed paraffin-embedded tumors with varying histopathological features. The two molecular layers exhibited modest correlation with each other, however, both datasets demonstrated stronger intertumoral heterogeneity compared to intratumoral heterogeneity. We identified potential markers and dysregulated pathways linked to tumors and normal adjacent tissues, and to the extent of immune infiltration, mucin and stroma content, including markers that behave distinctively in *ALK*-rearranged LADCs compared to non-*ALK*-driven LADCs. The extracellular matrix elements, particularly fibronectin 1, exhibited substantial variability within tumors.

In summary, the studies presented in this thesis generated new hypotheses for future lung cancer research, providing a list of biomarkers and dysregulated pathways with potential clinical relevance, and contributed to our understanding of cancer biology.

8. REFERENCES

1. World Health Organization. Cancer statistics [Internet]. 2022 [updated 2022 Feb 3; cited 2023 June 9]. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–249.

3. Bogos K, Kiss Z, Gálffy G, Tamási L, Ostoros G, Müller V, Urbán L, Bittner N, Sárosi V, Vastag A, Polányi Z, Nagy-Erdei Z, Vokó Z, Nagy B, Horváth K, Rokszin G, Abonyi-Tóth Z, Barcza Z, Moldvay J. Lung Cancer in Hungary. J Thorac Oncol. 2020;15(5):692–699.

4. Marino FZ, Bianco R, Accardo M, Ronchi A, Cozzolino I, Morgillo F, Rossi G, Franco R. Molecular heterogeneity in lung cancer: from mechanisms of origin to clinical implications. Int J Med Sci. 2019;16(7):981–989.

5. Haga Y, Minegishi Y, Ueda K. Frontiers in mass spectrometry–based clinical proteomics for cancer diagnosis and treatment. Cancer Sci. 2023;114(5):1783–1791.

6. Chen Y-J, Roumeliotis TI, Chang Y-H, Chen C-T, Han C-L, Lin M-H, Chen H-W, Chang G-C, Chang Y-L, Wu C-T, Lin M-W, Hsieh M-S, Wang Y-T, Chen Y-R, Jonassen I, Ghavidel FZ, Lin Z-S, Lin K-T, Chen C-W, Sheu P-Y, Hung C-T, Huang K-C, Yang H-C, Lin P-Y, Yen T-C, Lin Y-W, Wang J-H, Raghav L, Lin C-Y, Chen Y-S, Wu P-S, Lai C-T, Weng S-H, Su K-Y, Chang W-H, Tsai P-Y, Robles AI, Rodriguez H, Hsiao Y-J, Chang W-H, Sung T-Y, Chen J-S, Yu S-L, Choudhary JS, Chen H-Y, Yang P-C, Chen Y-J. Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. Cell. 2020;182(1):226–244.e17.

7. Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, Petralia F, Li Y, Liang W-W, Reva B, Krek A, Ji J, Song X, Liu W, Hong R, Yao L, Blumenberg L, Savage SR, Wendl MC, Wen B, Li K, Tang LC, MacMullan MA, Avanessian SC, Kane MH, Newton CJ, Cornwell M, Kothadia RB, Ma W, Yoo S, Mannan R, Vats P, Kumar-Sinha C, Kawaler EA, Omelchenko T, Colaprico A, Geffen Y, Maruvka YE, da Veiga Leprevost F, Wiznerowicz M, Gümüş ZH, Veluswamy RR, Hostetter G, Heiman DI, Wyczalkowski MA, Hiltke T, Mesri M, Kinsinger CR, Boja ES, Omenn GS, Chinnaiyan AM, Rodriguez H, Li QK, Jewell SD, Thiagarajan M, Getz G, Zhang B,

Fenyö D, Ruggles KV, Cieslik MP, Robles AI, Clauser KR, Govindan R, Wang P, Nesvizhskii AI, Ding L, Mani DR, Carr SA, Webster A, Francis A, Charamut A, Paulovich AG, Perou AM, Godwin AK, Karnuta A, Marrero-Oliveras A, Hindenach B, Pruetz B, Kubisa B, Druker BJ, Birger C, Jones CD, Valley DR, Rohrer DC, Zhou DC, Chan DW, Chesla D, Clark DJ, Rykunov D, Tan D, Ponomareva EV, Duffy E, Burks EJ, Schadt EE, Bergstrom EJ, Fedorov ES, Malc E, Wilson GD, Chen H-Q, Krzystek HM, Liu H, Culpepper H, Sun H, Zhang H, Day J, Suh J, Whiteaker JR, Eschbacher J, McGee J, Ketchum KA, Rodland KD, Robinson K, Hoadley KA, Suzuki K, Um KS, Elburn K, Wang L-B, Chen L, Hannick L, Qi L, Sokoll LJ, Wojtyś M, Domagalski MJ, Gritsenko MA, Beasley MB, Monroe ME, Ellis MJ, Dyer M, Burke MC, Borucki M, Sun M-H, Roehrl MH, Birrer MJ, Noble M, Schnaubelt M, Vernon M, Chaikin M, Krotevich M, Khan M, Selvan ME, Roche N, Edwards NJ, Vatanian N, Potapova O, Grady P, McGarvey PB, Mieczkowski P, Hariharan P, Madan R, Thangudu RR, Smith RD, Welsh RJ, Zelt R, Mehra R, Matteotti R, Mareedu S, Payne SH, Cottingham S, Markey SP, Chugh S, Smith S, Tsang S, Cai S, Boca SM, Carter S, Gabriel S, De Young S, Stein SE, Shankar S, Krubit T, Liu T, Skelly T, Bauer T, Velvulou U, Ozbek U, Petyuk VA, Sovenko V, Bocik WE, Maggio WW, Chen X, Shi Y, Wu Y, Hu Y, Liao Y, Zhang Z, Shi Z. Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell. 2020;182(1):200–225.e35.

8. Lehtiö J, Arslan T, Siavelis I, Pan Y, Socciarelli F, Berkovska O, Umer HM, Mermelekas G, Pirmoradian M, Jönsson M, Brunnström H, Brustugun OT, Purohit KP, Cunningham R, Foroughi Asl H, Isaksson S, Arbajian E, Aine M, Karlsson A, Kotevska M, Gram Hansen C, Drageset Haakensen V, Helland Å, Tamborero D, Johansson HJ, Branca RM, Planck M, Staaf J, Orre LM. Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune-evasion mechanisms. Nat Cancer. 2021;2(11):1224–1242.

9. Satpathy S, Krug K, Jean Beltran PM, Savage SR, Petralia F, Kumar-Sinha C, Dou Y, Reva B, Kane MH, Avanessian SC, Vasaikar SV, Krek A, Lei JT, Jaehnig EJ, Omelchenko T, Geffen Y, Bergstrom EJ, Stathias V, Christianson KE, Heiman DI, Cieslik MP, Cao S, Song X, Ji J, Liu W, Li K, Wen B, Li Y, Gümüş ZH, Selvan ME, Soundararajan R, Visal TH, Raso MG, Parra ER, Babur Ö, Vats P, Anand S, Schraink T, Cornwell M, Rodrigues FM, Zhu H, Mo C-K, Zhang Y, da Veiga Leprevost F, Huang C,

Chinnaiyan AM, Wyczalkowski MA, Omenn GS, Newton CJ, Schurer S, Ruggles KV, Fenyö D, Jewell SD, Thiagarajan M, Mesri M, Rodriguez H, Mani SA, Udeshi ND, Getz G, Suh J, Li QK, Hostetter G, Paik PK, Dhanasekaran SM, Govindan R, Ding L, Robles AI, Clauser KR, Nesvizhskii AI, Wang P, Carr SA, Zhang B, Mani DR, Gillette MA, Consortium CPTA. A proteogenomic portrait of lung squamous cell carcinoma. Cell. 2021;184(16):4348–4371.e40.

10. Soltis AR, Bateman NW, Liu J, Nguyen T, Franks TJ, Zhang X, Dalgard CL, Viollet C, Somiari S, Yan C, Zeman K, Skinner WJ, Lee JSH, Pollard HB, Turner C, Petricoin EF, Meerzaman D, Conrads TP, Hu H, APOLLO Research Network, Shriver CD, Moskaluk CA, Browning RF, Wilkerson MD. Proteogenomic analysis of lung adenocarcinoma reveals tumor heterogeneity, survival determinants, and therapeutically relevant pathways. Cell Rep Med. 2022;3(11):100819.

11. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. Lancet. 2021;398(10299):535–554.

12. Schwendenwein A, Megyesfalvi Z, Barany N, Valko Z, Bugyik E, Lang C, Ferencz B, Paku S, Lantos A, Fillinger J, Rezeli M, Marko-Varga G, Bogos K, Galffy G, Renyi-Vamos F, Hoda MA, Klepetko W, Hoetzenecker K, Laszlo V, Dome B. Molecular profiles of small cell lung cancer subtypes: therapeutic implications. Mol Ther Oncolytics. 2021;20:470–483.

13. Raso MG, Bota-Rabassedas N, Wistuba II. Pathology and Classification of SCLC. Cancers (Basel). 2021;13(4):820.

14. Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. Nat Rev Dis Primers. 2021;7(1):3.

15. Taniguchi H, Sen T, Rudin CM. Targeted Therapies and Biomarkers in Small Cell Lung Cancer. Front Oncol. 2020;10:741.

16. Rudin CM, Poirier JT, Byers LA, Dive C, Dowlati A, George J, Heymach JV, Johnson JE, Lehman JM, MacPherson D, Massion PP, Minna JD, Oliver TG, Quaranta V, Sage J, Thomas RK, Vakoc CR, Gazdar AF. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat Rev Cancer. 2019;19(5):289–297.

17. Zhang W, Girard L, Zhang Y-A, Haruki T, Papari-Zareei M, Stastny V, Ghayee HK, Pacak K, Oliver TG, Minna JD, Gazdar AF. Small cell lung cancer tumors and

preclinical models display heterogeneity of neuroendocrine phenotypes. Transl Lung Cancer Res. 2018;7(1):32–49.

18. George J, Lim JS, Jang SJ, Cun Y, Ozretić L, Kong G, Leenders F, Lu X, Fernández-Cuesta L, Bosco G, Müller C, Dahmen I, Jahchan NS, Park K-S, Yang D, Karnezis AN, Vaka D, Torres A, Wang MS, Korbel JO, Menon R, Chun S-M, Kim D, Wilkerson M, Hayes N, Engelmann D, Pützer B, Bos M, Michels S, Vlasic I, Seidel D, Pinther B, Schaub P, Becker C, Altmüller J, Yokota J, Kohno T, Iwakawa R, Tsuta K, Noguchi M, Muley T, Hoffmann H, Schnabel PA, Petersen I, Chen Y, Soltermann A, Tischler V, Choi C-m, Kim Y-H, Massion PP, Zou Y, Jovanovic D, Kontic M, Wright GM, Russell PA, Solomon B, Koch I, Lindner M, Muscarella LA, la Torre A, Field JK, Jakopovic M, Knezevic J, Castaños-Vélez E, Roz L, Pastorino U, Brustugun O-T, Lund-Iversen M, Thunnissen E, Köhler J, Schuler M, Botling J, Sandelin M, Sanchez-Cespedes M, Salvesen HB, Achter V, Lang U, Bogus M, Schneider PM, Zander T, Ansén S, Hallek M, Wolf J, Vingron M, Yatabe Y, Travis WD, Nürnberg P, Reinhardt C, Perner S, Heukamp L, Büttner R, Haas SA, Brambilla E, Peifer M, Sage J, Thomas RK. Comprehensive genomic profiles of small cell lung cancer. Nature. 2015;524(7563):47–53.

19. Drapkin BJ, Rudin CM. Advances in Small-Cell Lung Cancer (SCLC) Translational Research. Cold Spring Harb Perspect Med. 2021;11(4):a038240.

20. Caeser R, Egger JV, Chavan S, Socci ND, Jones CB, Kombak FE, Asher M, Roehrl MH, Shah NS, Allaj V, Manoj P, Tischfield SE, Kulick A, Meneses M, Iacobuzio-Donahue CA, Lai WV, Bhanot U, Baine MK, Rekhtman N, Hollmann TJ, de Stanchina E, Poirier JT, Rudin CM, Sen T. Genomic and transcriptomic analysis of a library of small cell lung cancer patient-derived xenografts. Nat Commun. 2022;13(1):2144.

21. Baine MK, Hsieh M-S, Lai WV, Egger JV, Jungbluth AA, Daneshbod Y, Beras A, Spencer R, Lopardo J, Bodd F, Montecalvo J, Sauter JL, Chang JC, Buonocore DJ, Travis WD, Sen T, Poirier JT, Rudin CM, Rekhtman N. SCLC Subtypes Defined by ASCL1, NEUROD1, POU2F3, and YAP1: A Comprehensive Immunohistochemical and Histopathologic Characterization. J Thorac Oncol. 2020;15(12):1823-1835.

22. Gay CM, Stewart CA, Park EM, Diao L, Groves SM, Heeke S, Nabet BY, Fujimoto J, Solis LM, Lu W, Xi Y, Cardnell RJ, Wang Q, Fabbri G, Cargill KR, Vokes NI, Ramkumar K, Zhang B, Della Corte CM, Robson P, Swisher SG, Roth JA, Glisson

BS, Shames DS, Wistuba II, Wang J, Quaranta V, Minna J, Heymach JV, Byers LA. Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. Cancer Cell. 2021;39(3):346–360.e7.

23. Liu Q, Zhang J, Guo C, Wang M, Wang C, Yan Y, Sun L, Wang D, Zhang L, Yu H, Hou L, Wu C, Zhu Y, Jiang G, Zhu H, Zhou Y, Fang S, Zhang T, Hu L, Li J, Liu Y, Zhang H, Zhang B, Ding L, Robles AI, Rodriguez H, Gao D, Ji H, Zhou H, Zhang P. Proteogenomic characterization of small cell lung cancer identifies biological insights and subtype-specific therapeutic strategies. Cell. 2024;187(1):184–203.e28.

24. Jeong H-C, Kim G-I, Cho S-H, Lee K-H, Ko J-J, Yang J-H, Chung K-H. Proteomic analysis of human small cell lung cancer tissues: up-regulation of coactosin-like protein-1. J Proteome Res. 2011;10(1):269–276.

25. Byers LA, Wang J, Nilsson MB, Fujimoto J, Saintigny P, Yordy J, Giri U, Peyton M, Fan YH, Diao L, Masrorpour F, Shen L, Liu W, Duchemann B, Tumula P, Bhardwaj V, Welsh J, Weber S, Glisson BS, Kalhor N, Wistuba II, Girard L, Lippman SM, Mills GB, Coombes KR, Weinstein JN, Minna JD, Heymach JV. Proteomic profiling identifies dysregulated pathways in small cell lung cancer and novel therapeutic targets including PARP1. Cancer Discov. 2012;2(9):798–811.

26. Fujii K, Miyata Y, Takahashi I, Koizumi H, Saji H, Hoshikawa M, Takagi M, Nishimura T, Nakamura H. Differential Proteomic Analysis between Small Cell Lung Carcinoma (SCLC) and Pulmonary Carcinoid Tumors Reveals Molecular Signatures for Malignancy in Lung Cancer. Proteomics Clin Appl. 2018;12(6):e1800015.

27. Gonçalves E, Poulos RC, Cai Z, Barthorpe S, Manda SS, Lucas N, Beck A, Bucio-Noble D, Dausmann M, Hall C, Hecker M, Koh J, Lightfoot H, Mahboob S, Mali I, Morris J, Richardson L, Seneviratne AJ, Shepherd R, Sykes E, Thomas F, Valentini S, Williams SG, Wu Y, Xavier D, MacKenzie KL, Hains PG, Tully B, Robinson PJ, Zhong Q, Garnett MJ, Reddel RR. Pan-cancer proteomic map of 949 human cell lines. Cancer Cell. 2022;40(8):835–849.e8.

28. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014;14(8):535–546.

29. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, Beer DG, Powell CA, Riely GJ, Van Schil PE, Garg K, Austin JHM, Asamura H, Rusch

VW, Hirsch FR, Scagliotti G, Mitsudomi T, Huber RM, Ishikawa Y, Jett J, Sanchez-Cespedes M, Sculier J-P, Takahashi T, Tsuboi M, Vansteenkiste J, Wistuba I, Yang P-C, Aberle D, Brambilla C, Flieder D, Franklin W, Gazdar A, Gould M, Hasleton P, Henderson D, Johnson B, Johnson D, Kerr K, Kuriyama K, Lee JS, Miller VA, Petersen I, Roggli V, Rosell R, Saijo N, Thunnissen E, Tsao M, Yankelewitz D. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol. 2011;6(2):244–285.

30. Karasaki T, Moore DA, Veeriah S, Naceur-Lombardelli C, Toncheva A, Magno N, Ward S, Bakir MA, Watkins TBK, Grigoriadis K, Huebner A, Hill MS, Frankell AM, Abbosh C, Puttick C, Zhai H, Gimeno-Valiente F, Saghafinia S, Kanu N, Dietzen M, Pich O, Lim EL, Martínez-Ruiz C, Black JRM, Biswas D, Campbell BB, Lee C, Colliver E, Enfield KSS, Hessey S, Hiley CT, Zaccaria S, Litchfield K, Birkbak NJ, Cadieux EL, Demeulemeester J, Van Loo P, Adusumilli PS, Tan KS, Cheema W, Sanchez-Vega F, Jones DR, Rekhtman N, Travis WD, Hackshaw A, Marafioti T, Salgado R, Le Quesne J, Nicholson AG, Consortium T, McGranahan N, Swanton C, Jamal-Hanjani M. Evolutionary characterization of lung adenocarcinoma morphology in TRACERx. Nat Med. 2023;29(4):833–845.

31. Bremnes RM, Busund LT, Kilvær TL, Andersen S, Richardsen E, Paulsen EE, Hald S, Khanehkenari MR, Cooper WA, Kao SC, Dønnem T. The Role of Tumor-Infiltrating Lymphocytes in Development, Progression, and Prognosis of Non-Small Cell Lung Cancer. J Thorac Oncol. 2016;11(6):789–800.

32. Lakshmanan I, Ponnusamy MP, Macha MA, Haridas D, Majhi PD, Kaur S, Jain M, Batra SK, Ganti AK. Mucins in Lung Cancer: Diagnostic, Prognostic, and Therapeutic Implications. J Thorac Oncol. 2015;10(1):19–27.

33. Valkenburg KC, de Groot AE, Pienta KJ. Targeting the tumour stroma to improve cancer therapy. Nat Rev Clin Oncol. 2018;15(6):366–381.

34. Saito M, Suzuki H, Kono K, Takenoshita S, Kohno T. Treatment of lung adenocarcinoma by molecular-targeted therapy and immunotherapy. Surg Today. 2018;48(1):1–8.

35. Planchard D, Popat S, Kerr K, Novello S, Smit EF, Faivre-Finn C, Mok TS, Reck M, Schil PEV, Hellmann MD, Peters S. Metastatic non-small cell lung cancer: ESMO

Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29:iv192–iv237.

36. Morris SW, Kirstein MN, Valentine MB, Dittmer KG, Shapiro DN, Saltman DL, Look AT. Fusion of a Kinase Gene, ALK, to a Nucleolar Protein Gene, NPM, in Non-Hodgkin's Lymphoma. Science. 1994;263(5151):1281–1284.

37. Peng L, Zhu L, Sun Y, Stebbing J, Selvaggi G, Zhang Y, Yu Z. Targeting ALK Rearrangements in NSCLC: Current State of the Art. Front Oncol. 2022;12(863461).

38. Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. Cancer Cell. 2020;37(4):471–484.

39. Marino FZ, Liguori G, Aquino G, Mantia EL, Bosari S, Ferrero S, Rosso L, Gaudioso G, Rosa ND, Scrima M, Martucci N, Rocca AL, Normanno N, Morabito A, Rocco G, Botti G, Franco R. Intratumor Heterogeneity of ALK-Rearrangements and Homogeneity of EGFR-Mutations in Mixed Lung Adenocarcinoma. PLoS One. 2015;10(9):e0139264.

40. Larroquette M, Guegan J-P, Besse B, Cousin S, Brunet M, Moulec SL, Loarer FL, Rey C, Soria J-C, Barlesi F, Bessede A, Scoazec J-Y, Soubeyran I, Italiano A. Spatial transcriptomics of macrophage infiltration in non-small cell lung cancer reveals determinants of sensitivity and resistance to anti-PD1/PD-L1 antibodies. J Immunother Cancer. 2022;10(5):e003890.

41. Monkman J, Taheri T, Ebrahimi Warkiani M, O'Leary C, Ladwa R, Richard D, O'Byrne K, Kulasinghe A. High-Plex and High-Throughput Digital Spatial Profiling of Non-Small-Cell Lung Cancer (NSCLC). Cancers (Basel). 2020;12(12):3551.

42. Moutafi MK, Molero M, Morilla SM, Baena J, Vathiotis IA, Gavrielatou N, Castro-Labrador L, Garibay GRd, Adradas V, Orive D, Valencia K, Calvo A, Montuenga LM, Aix SP, Schalper KA, Herbst RS, Paz-Ares L, Rimm DL, Zugazagoitia J. Spatially resolved proteomic profiling identifies tumor cell CD44 as a biomarker associated with sensitivity to PD-1 axis blockade in advanced non-small-cell lung cancer. J Immunother Cancer. 2022;10(8):e004757.

43. Zugazagoitia J, Gupta S, Liu Y, Fuhrman K, Gettinger S, Herbst RS, Schalper KA, Rimm DL. Biomarkers Associated with Beneficial PD-1 Checkpoint Blockade in Non–Small Cell Lung Cancer (NSCLC) Identified Using High-Plex Digital Spatial Profiling. Clin Cancer Res. 2020;26(16):4360–4368.

44. Li Y, Dou Y, Da Veiga Leprevost F, Geffen Y, Calinawan AP, Aguet F, Akiyama Y, Anand S, Birger C, Cao S, Chaudhary R, Chilappagari P, Cieslik M, Colaprico A, Zhou DC, Day C, Domagalski MJ, Esai Selvan M, Fenyö D, Foltz SM, Francis A, Gonzalez-Robles T, Gümüş ZH, Heiman D, Holck M, Hong R, Hu Y, Jaehnig EJ, Ji J, Jiang W, Katsnelson L, Ketchum KA, Klein RJ, Lei JT, Liang W-W, Liao Y, Lindgren CM, Ma W, Ma L, MacCoss MJ, Martins Rodrigues F, McKerrow W, Nguyen N, Oldroyd R, Pilozzi A, Pugliese P, Reva B, Rudnick P, Ruggles KV, Rykunov D, Savage SR, Schnaubelt M, Schraink T, Shi Z, Singhal D, Song X, Storrs E, Terekhanova NV, Thangudu RR, Thiagarajan M, Wang L-B, Wang JM, Wang Y, Wen B, Wu Y, Wyczalkowski MA, Xin Y, Yao L, Yi X, Zhang H, Zhang Q, Zuhl M, Getz G, Ding L, Nesvizhskii AI, Wang P, Robles AI, Zhang B, Payne SH, Consortium CPTA. Proteogenomic data and resources for pan-cancer analysis. Cancer Cell. 2023;41(8):1397–1406.

45. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA, Ge Y, Gunawardena J, Hendrickson RC, Hergenrother PJ, Huber CG, Ivanov AR, Jensen ON, Jewett MC, Kelleher NL, Kiessling LL, Krogan NJ, Larsen MR, Loo JA, Ogorzalek Loo RR, Lundberg E, MacCoss MJ, Mallick P, Mootha VK, Mrksich M, Muir TW, Patrie SM, Pesavento JJ, Pitteri SJ, Rodriguez H, Saghatelian A, Sandoval W, Schlüter H, Sechi S, Slavoff SA, Smith LM, Snyder MP, Thomas PM, Uhlén M, Van Eyk JE, Vidal M, Walt DR, White FM, Williams ER, Wohlschlager T, Wysocki VH, Yates NA, Young NL, Zhang B. How many human proteoforms are there? Nat Chem Biol. 2018;14(3):206–214.

46. Smith LM, Kelleher NL, Proteomics CfTD. Proteoform: a single term describing protein complexity. Nat Methods. 2013;10(3):186–187.

47. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F, Kawaler E, Mundt F, Krug K, Tu Z, Lei JT, Gatza ML, Wilkerson M, Perou CM, Yellapantula V, Huang K-l, Lin C, McLellan MD, Yan P, Davies SR, Townsend RR, Skates SJ, Wang J, Zhang B, Kinsinger CR, Mesri M, Rodriguez H, Ding L, Paulovich AG, Fenyö D, Ellis MJ, Carr SA, CPTAC N. Proteogenomics connects somatic mutations to signalling in breast cancer. Nature. 2016;534(7605):55–62.

48. Sinha A, Huang V, Livingstone J, Wang J, Fox NS, Kurganovs N, Ignatchenko V, Fritsch K, Donmez N, Heisler LE, Shiah Y-J, Yao CQ, Alfaro JA, Volik S, Lapuk A, Fraser M, Kron K, Murison A, Lupien M, Sahinalp C, Collins CC, Tetu B, Masoomian M, Berman DM, van der Kwast T, Bristow RG, Kislinger T, Boutros PC. The Proteogenomic Landscape of Curable Prostate Cancer. Cancer Cell. 2019;35(3):414–427.e6.

49. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, Davies SR, Wang S, Wang P, Kinsinger CR, Rivers RC, Rodriguez H, Townsend RR, Ellis MJC, Carr SA, Tabb DL, Coffey RJ, Slebos RJC, Liebler DC, CPTAC N. Proteogenomic characterization of human colon and rectal cancer. Nature. 2014;513(7518):382–387.

50. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou J-Y, Petyuk VA, Chen L, Ray D, Sun S, Yang F, Chen L, Wang J, Shah P, Cha SW, Aiyetan P, Woo S, Tian Y, Gritsenko MA, Clauss TR, Choi C, Monroe ME, Thomas S, Nie S, Wu C, Moore RJ, Yu K-H, Tabb DL, Fenyö D, Bafna V, Wang Y, Rodriguez H, Boja ES, Hiltke T, Rivers RC, Sokoll L, Zhu H, Shih I-M, Cope L, Pandey A, Zhang B, Snyder MP, Levine DA, Smith RD, Chan DW, Rodland KD, Investigators C. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell. 2016;166(3):755–765.

51. Brown KA, Melby JA, Roberts DS, Ge Y. Top-down proteomics: challenges, innovations, and applications in basic and clinical research. Expert Rev Proteomics. 2020;17(10):719–733.

52. Miller RM, Smith LM. Overview and considerations in bottom-up proteomics. Analyst. 2023;148(3):475–486.

53. Muntel J, Gandhi T, Verbeke L, Bernhardt OM, Treiber T, Bruderer R, Reiter L. Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. Mol Omics. 2019;15(5):348–360.

54. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature. 2016;537(7620):347–355.

55. Gil J, Betancourt LH, Pla I, Sanchez A, Appelqvist R, Miliotis T, Kuras M, Oskolas H, Kim Y, Horvath Z, Eriksson J, Berge E, Burestedt E, Jönsson G, Baldetorp B, Ingvar C, Olsson H, Lundgren L, Horvatovich P, Murillo JR, Sugihara Y, Welinder C,

Wieslander E, Lee B, Lindberg H, Pawłowski K, Kwon HJ, Doma V, Timar J, Karpati S, Szasz AM, Németh IB, Nishimura T, Corthals G, Rezeli M, Knudsen B, Malm J, Marko-Varga G. Clinical protein science in translational medicine targeting malignant melanoma. Cell Biol Toxicol. 2019;35(4):293–332.

56. Kitata RB, Yang J-C, Chen Y-J. Advances in data-independent acquisition mass spectrometry towards comprehensive digital proteome landscape. Mass Spectrom Rev. 2023;42(6):2324–2348.

57. Teo GC, Polasky DA, Yu F, Nesvizhskii AI. Fast Deisotoping Algorithm and Its Implementation in the MSFragger Search Engine. J Proteome Res. 2021;20(1):498–505.

58. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367–1372.

59. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966–968.

60. Solntsev SK, Shortreed MR, Frey BL, Smith LM. Enhanced Global Post-translational Modification Discovery with MetaMorpheus. J Proteome Res. 2018;17(5):1844–1851.

61. Muth T, Renard BY. Evaluating de novo sequencing in proteomics: already an accurate alternative to database-driven peptide identification? Brief Bioinform. 2018;19(5):954–970.

62. Tabb DL. The SEQUEST family tree. J Am Soc Mass Spectrom. 2015;26(11):1814–1819.

63. Griss J. Spectral library searching in proteomics. Proteomics. 2016;16(5):729–740.

64. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. J Proteomics. 2010;73(11):2092–2123.

65. Taverna D, Gaspari M. A critical comparison of three MS-based approaches for quantitative proteomics analysis. J Mass Spectrom. 2021;56(1):e4669.

66. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem. 2007;389(4):1017-1031.

67. Välikangas T, Suomi T, Elo LL. A systematic evaluation of normalization methods in quantitative label-free proteomics. Brief Bioinform. 2018;19(1):1–11.

68. Čuklina J, Lee CH, Williams EG, Sajic T, Collins BC, Rodríguez Martínez M, Sharma VS, Wendt F, Goetze S, Keele GR, Wollscheid B, Aebersold R, Pedrioli PGA. Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. Mol Syst Biol. 2021;17(8):e10240.

69. Lazar C, Gatto L, Ferro M, Bruley C, Burger T. Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. J Proteome Res. 2016;15(4):1116–1125.

70. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. Bioinformatics. 2001;17(6):520–525.

71. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13(9):731–740.

72. Kohler D, Staniak M, Tsai T-H, Huang T, Shulman N, Bernhardt OM, MacLean BX, Nesvizhskii AI, Reiter L, Sabido E, Choi M, Vitek O. MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale. J Proteome Res. 2023;22(5):1466–1482.

73. Berrar D, Granzow M, Dubitzky W. Introduction to Genomic and Proteomic Data Analysis. In: Dubitzky W, Granzow M, Berrar D, editors. Fundamentals of Data Mining in Genomics and Proteomics. Boston, MA: Springer US; 2007. p. 1–37.

74. Menyhart O, Weltz B, Győrffy B. MultipleTesting.com: A tool for life science researchers for multiple hypothesis testing correction. PLoS One. 2021;16(6):e0245824.

75. Wu X, Hasan MA, Chen JY. Pathway and network analysis in proteomics. J Theor Biol. 2014;362:44–52.

76. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13.

77. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–15550.

78. Errington TM, Denis A, Perfito N, Iorns E, Nosek BA. Challenges for assessing replicability in preclinical cancer biology. Elife. 2021;10:e67995.

79. Szeitz B, Megyesfalvi Z, Woldmar N, Valkó Z, Schwendenwein A, Bárány N, Paku S, László V, Kiss H, Bugyik E, Lang C, Szász AM, Pizzatti L, Bogos K, Hoda MA, Hoetzenecker K, Marko-Varga G, Horvatovich P, 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.

80. Szeitz B, Glasz T, Herold Z, Tóth G, Balbisi M, Fillinger J, Horváth S, Mohácsi R, Kwon HJ, Moldvay J, Turiák L, 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.

81. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, Hamilton PW. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7(1):16878.

82. Kuras M, Woldmar N, Kim Y, Hefner M, Malm J, Moldvay J, Döme B, Fillinger J, Pizzatti L, Gil J, Marko-Varga G, Rezeli M. Proteomic Workflows for High-Quality Quantitative Proteome and Post-Translational Modification Analysis of Clinically Relevant Samples from Formalin-Fixed Paraffin-Embedded Archives. J Proteome Res. 2021;20(1):1027–1039.

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

84. Bugyi F, Tóth G, Kovács KB, Drahos L, Turiák L. Comparison of solid-phase extraction methods for efficient purification of phosphopeptides with low sample amounts. J Chromatogr A. 2022;1685:463597.

85. Turiák L, Ozohanics O, Tóth G, Ács A, Révész Á, Vékey K, Telekes A, Drahos L. High sensitivity proteomics of prostate cancer tissue microarrays to discriminate between healthy and cancerous tissue. J Proteomics. 2019;197:82–91.

86. Uhlén M, Karlsson MJ, Hober A, Svensson A-S, Scheffel J, Kotol D, Zhong W, Tebani A, Strandberg L, Edfors F, Sjöstedt E, Mulder J, Mardinoglu A, Berling A, Ekblad S, Dannemeyer M, Kanje S, Rockberg J, Lundqvist M, Malm M, Volk A-L, Nilsson P, Månberg A, Dodig-Crnkovic T, Pin E, Zwahlen M, Oksvold P, von Feilitzen K, Häussler RS, Hong M-G, Lindskog C, Ponten F, Katona B, Vuu J, Lindström E, Nielsen J, Robinson J, Ayoglu B, Mahdessian D, Sullivan D, Thul P, Danielsson F, Stadler C, Lundberg E, Bergström G, Gummesson A, Voldborg BG, Tegel H, Hober S, Forsström B, Schwenk JM, Fagerberg L, Sivertsson Å. The human secretome. Sci Signal. 2019;12(609):eaaz0274.

87. Chen G, Chen J, Liu H, Chen S, Zhang Y, Li P, Thierry-Mieg D, Thierry-Mieg J, Mattes W, Ning B, Shi T. Comprehensive Identification and Characterization of Human Secretome Based on Integrative Proteomic and Transcriptomic Data. Front Cell Dev Biol. 2019;7:299.

88. Meinken J, Walker G, Cooper CR, Min XJ. MetazSecKB: the human and animal secretome and subcellular proteome knowledgebase. Database (Oxford). 2015;2015:bav077.

89. Consortium U. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480–D489.

90. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074–D1082.

91. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, Ramaswamy S, Futreal PA, Haber DA, Stratton MR, Benes C, McDermott U, Garnett MJ. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013;41(Database issue):D955–961.

92. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N.

The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discov. 2012;2(5):401–404.

93. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1.

94. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, Barretina J, Gelfand ET, Bielski CM, Li H, Hu K, Andreev-Drakhlin AY, Kim J, Hess JM, Haas BJ, Aguet F, Weir BA, Rothberg MV, Paolella BR, Lawrence MS, Akbani R, Lu Y, Tiv HL, Gokhale PC, de Weck A, Mansour AA, Oh C, Shih J, Hadi K, Rosen Y, Bistline J, Venkatesan K, Reddy A, Sonkin D, Liu M, Lehar J, Korn JM, Porter DA, Jones MD, Golji J, Caponigro G, Taylor JE, Dunning CM, Creech AL, Warren AC, McFarland JM, Zamanighomi M, Kauffmann A, Stransky N, Imielinski M, Maruvka YE, Cherniack AD, Tsherniak A, Vazquez F, Jaffe JD, Lane AA, Weinstock DM, Johannessen CM, Morrissey MP, Stegmeier F, Schlegel R, Hahn WC, Getz G, Mills GB, Boehm JS, Golub TR, Garraway LA, Sellers WR. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569(7757):503–508.

95. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

96. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Fröhling S, Chan EM, Sos ML, Michel K, Mermel C, Silver SJ, Weir BA, Reiling JH, Sheng Q, Gupta PB, Wadlow RC, Le H, Hoersch S, Wittner BS, Ramaswamy S, Livingston DM, Sabatini DM, Meyerson M, Thomas RK, Lander ES, Mesirov JP, Root DE, Gilliland DG, Jacks T, Hahn WC. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462(7269):108–112.

97. Kohn KW, Zeeberg BM, Reinhold WC, Pommier Y. Gene expression correlations in human cancer cell lines define molecular interaction networks for epithelial phenotype. PLoS One. 2014;9(6):e99269.

98. Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, Davis MJ. Single sample scoring of molecular phenotypes. BMC Bioinformatics. 2018;19(1):404.

99. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417–425.

100. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.

101. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, Griss J, Sevilla C, Matthews L, Gong C, Deng C, Varusai T, Ragueneau E, Haider Y, May B, Shamovsky V, Weiser J, Brunson T, Sanati N, Beckman L, Shao X, Fabregat A, Sidiropoulos K, Murillo J, Viteri G, Cook J, Shorser S, Bader G, Demir E, Sander C, Haw R, Wu G, Stein L, Hermjakob H, D'Eustachio P. The Reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50(D1):D687–D692.

102. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739–1740.

103. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847–2849.

104. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572–1573.

105. Rivellese F, Surace AEA, Goldmann K, Sciacca E, Çubuk C, Giorli G, John CR, Nerviani A, Fossati-Jimack L, Thorborn G, Ahmed M, Prediletto E, Church SE, Hudson BM, Warren SE, McKeigue PM, Humby F, Bombardieri M, Barnes MR, Lewis MJ, Pitzalis C. Rituximab versus tocilizumab in rheumatoid arthritis: synovial biopsy-based biomarker analysis of the phase 4 R4RA randomized trial. Nat Med. 2022;28(6):1256–1268.

106. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS. 2012;16(5):284–287.

107. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016;12(2):477–479.

108. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene

ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–29.

109. Consortium GO. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021;49(D1):D325–D334.

110. Megyesfalvi Z, Barany N, Lantos A, Valko Z, Pipek O, Lang C, Schwendenwein A, Oberndorfer F, Paku S, Ferencz B, Dezso K, Fillinger J, Lohinai Z, Moldvay J, Galffy G, Szeitz B, Rezeli M, Rivard C, Hirsch FR, Brcic L, Popper H, Kern I, Kovacevic M, Skarda J, Mittak M, Marko-Varga G, Bogos K, Renyi-Vamos F, Hoda MA, Klikovits T, Hoetzenecker K, Schelch K, 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):674–686.

111. Carney DN, Gazdar AF, Bepler G, Guccion JG, Marangos PJ, Moody TW, Zweig MH, Minna JD. Establishment and identification of small cell lung cancer cell lines having classic and variant features. Cancer Res. 1985;45(6):2913–2923.

112. Gazdar AF, Carney DN, Nau MM, Minna JD. Characterization of variant subclasses of cell lines derived from small cell lung cancer having distinctive biochemical, morphological, and growth properties. Cancer Res. 1985;45(6):2924–2930.

113. Baine MK, Febres-Aldana CA, Chang JC, Jungbluth AA, Sethi S, Antonescu CR, Travis WD, Hsieh M-S, Roh MS, Homer RJ, Ladanyi M, Egger JV, Lai WV, Rudin CM, Rekhtman N. POU2F3 in SCLC: Clinicopathologic and Genomic Analysis With a Focus on Its Diagnostic Utility in Neuroendocrine-Low SCLC. J Thorac Oncol. 2022;17(9):1109–1121.

114. Tlemsani C, Pongor L, Elloumi F, Girard L, Huffman KE, Roper N, Varma S, Luna A, Rajapakse VN, Sebastian R, Kohn KW, Krushkal J, Aladjem MI, Teicher BA, Meltzer PS, Reinhold WC, Minna JD, Thomas A, Pommier Y. SCLC-CellMiner: A Resource for Small Cell Lung Cancer Cell Line Genomics and Pharmacology Based on Genomic Signatures. Cell Rep. 2020;33(3):108296.

115. Borromeo MD, Savage TK, Kollipara RK, He M, Augustyn A, Osborne JK, Girard L, Minna JD, Gazdar AF, Cobb MH, Johnson JE. ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. Cell Rep. 2016;16(5):1259–1272.

116. Sethi T, Rintoul RC, Moore SM, MacKinnon AC, Salter D, Choo C, Chilvers ER, Dransfield I, Donnelly SC, Strieter R, Haslett C. Extracellular matrix proteins protect small cell lung cancer cells against apoptosis: a mechanism for small cell lung cancer growth and drug resistance in vivo. Nat Med. 1999;5(6):662–668.

117. Zhao P, Sun X, Li H, Liu Y, Cui Y, Tian L, Cheng Y. c-Myc Targets HDAC3 to Suppress NKG2DL Expression and Innate Immune Response in N-Type SCLC through Histone Deacetylation. Cancers (Basel). 2022;14(3):457.

118. Chan JM, Quintanal-Villalonga Á, Gao VR, Xie Y, Allaj V, Chaudhary O, Masilionis I, Egger J, Chow A, Walle T, Mattar M, Yarlagadda DVK, Wang JL, Uddin F, Offin M, Ciampricotti M, Qeriqi B, Bahr A, de Stanchina E, Bhanot UK, Lai WV, Bott MJ, Jones DR, Ruiz A, Baine MK, Li Y, Rekhtman N, Poirier JT, Nawy T, Sen T, Mazutis L, Hollmann TJ, Pe'er D, Rudin CM. Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. Cancer Cell. 2021;39(11):1479–1496.e18.

119. Solta A, Boettiger K, Kovács I, Lang C, Megyesfalvi Z, Ferk F, Mišík M, Hoetzenecker K, Aigner C, Kowol CR, Knasmueller S, Grusch M, Szeitz B, Rezeli M, 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):4644–4659.

120. Guillemot F, Hassan BA. Beyond proneural: emerging functions and regulations of proneural proteins. Curr Opin Neurobiol. 2017;42:93–101.

121. Liao Y, Yin G, Wang X, Zhong P, Fan X, Huang C. Identification of candidate genes associated with the pathogenesis of small cell lung cancer via integrated bioinformatics analysis. Oncol Lett. 2019;18(4):3723–3733.

122. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol. 2014;15(3):178–196.

123. Hwang J-J, Choi S-Y, Koh J-Y. The role of NADPH oxidase, neuronal nitric oxide synthase and poly(ADP ribose) polymerase in oxidative neuronal death induced in cortical cultures by brain-derived neurotrophic factor and neurotrophin-4/5. J Neurochem. 2002;82(4):894–902.

124. Dupont S, Morsut L, Aragona M, Enzo E, Giulitti S, Cordenonsi M, Zanconato F, Le Digabel J, Forcato M, Bicciato S, Elvassore N, Piccolo S. Role of YAP/TAZ in mechanotransduction. Nature. 2011;474(7350):179–183.

125. Caeser R, Hulton C, Costa E, Durani V, Little M, Chen X, Tischfield SE, Asher M, Kombak FE, Chavan SS, Shah NS, Ciampricotti M, de Stanchina E, Poirier JT, Rudin CM, Sen T. MAPK pathway activation selectively inhibits ASCL1-driven small cell lung cancer. iScience. 2021;24(11):103224.

126. Gazdar AF, Gao B, Minna JD. Lung cancer cell lines: Useless artifacts or invaluable tools for medical science? Lung Cancer. 2010;68(3):309–318.

127. van Staveren WCG, Solís DYW, Hébrant A, Detours V, Dumont JE, Maenhaut C. Human cancer cell lines: Experimental models for cancer cells in situ? For cancer stem cells? Biochim Biophys Acta. 2009;1795(2):92–103.

128. Moreira AL, Ocampo PSS, Xia Y, Zhong H, Russell PA, Minami Y, Cooper WA, Yoshida A, Bubendorf L, Papotti M, Pelosi G, Lopez-Rios F, Kunitoki K, Ferrari-Light D, Sholl LM, Beasley MB, Borczuk A, Botling J, Brambilla E, Chen G, Chou T-Y, Chung J-H, Dacic S, Jain D, Hirsch FR, Hwang D, Lantuejoul S, Lin D, Longshore JW, Motoi N, Noguchi M, Poleri C, Rekhtman N, Tsao M-S, Thunnissen E, Travis WD, Yatabe Y, Roden AC, Daigneault JB, Wistuba II, Kerr KM, Pass H, Nicholson AG, Mino-Kenudson M. A Grading System for Invasive Pulmonary Adenocarcinoma: A Proposal From the International Association for the Study of Lung Cancer Pathology Committee. J Thorac Oncol. 2020;15(10):1599–1610.

129. Sharpnack MF, Ranbaduge N, Srivastava A, Cerciello F, Codreanu SG, Liebler DC, Mascaux C, Miles WO, Morris R, McDermott JE, Sharpnack JL, Amann J, Maher CA, Machiraju R, Wysocki VH, Govindan R, Mallick P, Coombes KR, Huang K, Carbone DP. Proteogenomic Analysis of Surgically Resected Lung Adenocarcinoma. J Thorac Oncol. 2018;13(10):1519–1529.

130. Stewart PA, Parapatics K, Welsh EA, Müller AC, Cao H, Fang B, Koomen JM, Eschrich SA, Bennett KL, Haura EB. A Pilot Proteogenomic Study with Data Integration Identifies MCT1 and GLUT1 as Prognostic Markers in Lung Adenocarcinoma. PLoS One. 2015;10(11):e0142162.

131. Xu J-Y, Zhang C, Wang X, Zhai L, Ma Y, Mao Y, Qian K, Sun C, Liu Z, Jiang S, Wang M, Feng L, Zhao L, Liu P, Wang B, Zhao X, Xie H, Yang X, Zhao L, Chang Y,

Jia J, Wang X, Zhang Y, Wang Y, Yang Y, Wu Z, Yang L, Liu B, Zhao T, Ren S, Sun A, Zhao Y, Ying W, Wang F, Wang G, Zhang Y, Cheng S, Qin J, Qian X, Wang Y, Li J, He F, Xiao T, Tan M. Integrative Proteomic Characterization of Human Lung Adenocarcinoma. Cell. 2020;182(1):245–261.e17.

132. Ruggles KV, Krug K, Wang X, Clauser KR, Wang J, Payne SH, Fenyö D, Zhang B, Mani DR. Methods, Tools and Current Perspectives in Proteogenomics. Mol Cell Proteomics. 2017;16(6):959–981.

133. Buccitelli C, Selbach M. mRNAs, proteins and the emerging principles of gene expression control. Nat Rev Genet. 2020;21(10):630–644.

134. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell. 2016;165(3):535–550.

135. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13(4):227–232.

136. Wei R, Qiu H, Xu J, Mo J, Liu Y, Gui Y, Huang G, Zhang S, Yao H, Huang X, Gan Z. Expression and prognostic potential of GPX1 in human cancers based on data mining. Ann Transl Med. 2020;8(4):124.

137. Zhao Y, Wang H, Zhou J, Shao Q. Glutathione Peroxidase GPX1 and Its Dichotomous Roles in Cancer. Cancers (Basel). 2022;14(10):2560.

138. Chen B, Shen Z, Wu D, Xie X, Xu X, Lv L, Dai H, Chen J, Gan X. Glutathione Peroxidase 1 Promotes NSCLC Resistance to Cisplatin via ROS-Induced Activation of PI3K/AKT Pathway. Biomed Res Int. 2019;2019:e7640547.

139. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science. 2009;324(5930):1029–1033.

140. Alhasan B, Mikeladze M, Guzhova I, Margulis B. Autophagy, molecular chaperones, and unfolded protein response as promoters of tumor recurrence. Cancer Metastasis Rev. 2023;42(1):217–254.

141. Pandolfi PP. Aberrant mRNA translation in cancer pathogenesis: an old concept revisited comes finally of age. Oncogene. 2004;23(18):3134–3137.

142. Parker AL, Cox TR. The Role of the ECM in Lung Cancer Dormancy and Outgrowth. Front Oncol. 2020;10:1766.

143. Du Z, Lovly CM. Mechanisms of receptor tyrosine kinase activation in cancer. Mol Cancer. 2018;17(1):58.

144. Schneider JL, Lin JJ, Shaw AT. ALK-positive lung cancer: a moving target. Nat Cancer. 2023;4(3):330–343.

145. Huang H. Anaplastic Lymphoma Kinase (ALK) Receptor Tyrosine Kinase: A Catalytic Receptor with Many Faces. Int J Mol Sci. 2018;19(11):3448.

146. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial-Mesenchymal Transitions in Development and Disease. Cell. 2009;139(5):871–890.

147. Lin JJ, Riely GJ, Shaw AT. Targeting ALK: Precision Medicine Takes on Drug Resistance. Cancer Discov. 2017;7(2):137–155.

148. Jena MK, Janjanam J. Role of extracellular matrix in breast cancer development: a brief update. F1000Res. 2018;7:274.

149. Hernández-Elvira M, Sunnerhagen P. Post-transcriptional regulation during stress. FEMS Yeast Res. 2022;22(1):foac025.

150. Lin T-C, Yang C-H, Cheng L-H, Chang W-T, Lin Y-R, Cheng H-C. Fibronectin in Cancer: Friend or Foe. Cells. 2020;9(1):27.

151. Di Modugno F, Spada S, Palermo B, Visca P, Iapicca P, Di Carlo A, Antoniani B, Sperduti I, Di Benedetto A, Terrenato I, Mottolese M, Gandolfi F, Facciolo F, Chen EI, Schwartz MA, Santoni A, Bissell MJ, Nisticò P. hMENA isoforms impact NSCLC patient outcome through fibronectin/β1 integrin axis. Oncogene. 2018;37(42):5605–5617.

152. Han H-j, Sung JY, Kim S-H, Yun U-J, Kim H, Jang E-J, Yoo H-E, Hong EK, Goh S-H, Moon A, Lee J-S, Ye S-K, Shim J, Kim Y-N. Fibronectin regulates anoikis resistance via cell aggregate formation. Cancer Lett. 2021;508:59–72.

153. Nishino T, Ishida T, Oka T, Yasumoto K, Sugimachi K. Distribution of fibronectin in adenocarcinoma of the lung: Classification and prognostic significance. J Surg Oncol. 1990;43(2):94–100.

154. Zhou M, Kong Y, Wang X, Li W, Chen S, Wang L, Wang C, Zhang Q. LC-MS/MS-Based Quantitative Proteomics Analysis of Different Stages of Non-Small-Cell Lung Cancer. Biomed Res Int. 2021;2021:e5561569.

155. Han J-Y, Kim HS, Lee SH, Park WS, Lee JY, Yoo NJ. Immunohistochemical expression of integrins and extracellular matrix proteins in non-small cell lung cancer: correlation with lymph node metastasis. Lung Cancer. 2003;41(1):65–70.

156. Sung WJ, Park K-S, Kwak SG, Hyun D-S, Jang JS, Park K-K. Epithelial-mesenchymal transition in patients of pulmonary adenocarcinoma: correlation with cancer stem cell markers and prognosis. Int J Clin Exp Pathol. 2015;8(8):8997–9009.

9. BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

9.1 Publications related to the subjects of the thesis

Szeitz B, Glasz T, Herold Z, Tóth G, Balbisi M, Fillinger J, Horváth S, Mohácsi R, Kwon HJ, Moldvay J, Turiák L, 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), SJR: D1 (Inorganic Chemistry, Spectroscopy)

Szeitz B*, Megyesfalvi Z*, Woldmar N, Valkó Z, Schwendenwein A, Bárány N, Paku S, László V, Kiss H, Bugyik E, Lang C, Szász AM, Pizzatti L, Bogos K, Hoda MA, Hoetzenecker K, Marko-Varga G, Horvatovich P, 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**, SJR: D1 (Medicine - miscellaneous)

**Shared first authors.*

9.2 Other publications

Solta A, Boettiger K, Kovács I, Lang C, Megyesfalvi Z, Ferk F, Mišík M, Hoetzenecker K, Aigner C, Kowol CR, Knasmueller S, Grusch M, **Szeitz B**, Rezeli M, 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):4644–4659. **IF: 11.5** (expected)

Tisza A, Klikovits T, Benej M, Torok S, **Szeitz B**, Valko Z, Hoda MA, Hegedus B, Bonta M, Nischkauer W, Hoetzenecker K, Limbeck A, Schelch K, Laszlo V, 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):3364–3374. **IF: 3.4** (expected)

Balbisi M, Sugár S, Schlosser G, **Szeitz B**, Fillinger J, Moldvay J, Drahos L, Szász AM, Tóth G, 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, Schwendenwein A, Kuras M, **Szeitz B**, Boettiger K, Tisza A, László V, Reiniger L, Bagó AG, Szállási Z, Moldvay J, Szász AM, Malm J, Horvatovich P, Pizzatti L, Domont GB, Rényi-Vámos F, Hoetzenecker K, Hoda MA, Marko-Varga G, Schelch K, Megyesfalvi Z, 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, Barany N, Lantos A, Valko Z, Pipek O, Lang C, Schwendenwein A, Oberndorfer F, Paku S, Ferencz B, Dezso K, Fillinger J, Lohinai Z, Moldvay J, Galffy G, **Szeitz B**, Rezeli M, Rivard C, Hirsch FR, Brcic L, Popper H, Kern I, Kovacevic M, Skarda J, Mittak M, Marko-Varga G, Bogos K, Renyi-Vamos F, Hoda MA, Klikovits T, Hoetzenecker K, Schelch K, 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):674–686. **IF: 7.3**

Szeitz B, Pipek O, Kulka J, Szundi C, Rusz O, Tőkés T, Szász AM, Kovács KA, Pesti A, Ben Arie TB, Gángó A, Fülöp Z, Drágus E, Vári-Kakas SA, 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**, Almeida NP, Domont G, Betancourt LH, Gil J, Marko-Varga M, Oskolas H, Jánosi Á J, Boyano-Adánez MDC, Kemény L, Baldetorp B, Malm J, Horvatovich P, Szász AM, 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***, Gil J, Rodriguez J, Palkovits M, Renner É, Hortobágyi T, Döme P, Nogueira FC, Marko-Varga G, 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, Moldvay J, Fillinger J, **Szeitz B**, Ferencz B, Boettiger K, Rezeli M, Bogos K, Renyi-Vamos F, Hoetzenecker K, Hegedus 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, Doma V, Çakır U, Sanchez A, Murillo JR, Kuras M, Parada IP, Sugihara Y, Appelqvist R, Wieslander E, Welinder C, Velasquez E, de Almeida NP, Woldmar N, Marko-Varga M, Pawłowski K, Eriksson J, **Szeitz B**, Baldetorp B, Ingvar C, Olsson H, Lundgren L, Lindberg H, Oskolas H, Lee B, Berge E, Sjögren M, Eriksson C, Kim D, Kwon HJ, Knudsen B, Rezeli M, Hong R, Horvatovich P, Miliotis T, Nishimura T, Kato H, Steinfelder E, Oppermann M, Miller K, Florindi F, Zhou Q, Domont GB, Pizzatti L, Nogueira FCS, Horvath P, Szadai L, Tímár J, Kárpáti S, Szász AM, Malm J, Fenyö D, Ekedahl H, 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, Sanchez A, Doma V, Kuras M, Murillo JR, Velasquez E, Çakır U, Kim Y, Sugihara Y, Parada IP, **Szeitz B**, Appelqvist R, Wieslander E, Welinder C, de Almeida NP, Woldmar N, Marko-Varga M, Eriksson J, Pawłowski K, Baldetorp B, Ingvar C, Olsson H, Lundgren L, Lindberg H, Oskolas H, Lee B, Berge E, Sjögren M, Eriksson C, Kim D, Kwon HJ, Knudsen B, Rezeli M, Malm J, Hong R, Horvath P, Szász AM, Tímár J, Kárpáti S, Horvatovich P, Miliotis T, Nishimura T, Kato H, Steinfelder E, Oppermann M, Miller K, Florindi F, Zhou Q, Domont GB, Pizzatti L, Nogueira FCS, Szadai L, Németh IB, Ekedahl H, Fenyö D, Marko-Varga G. The Human Melanoma Proteome Atlas-Complementing the melanoma transcriptome. Clin Transl Med. 2021;11(7):e451. **IF: 8.554**

DOI:10.14753/SE.2024.3031

10. ACKNOWLEDGEMENTS

The results presented in this thesis, as well as those from other works unrelated to the thesis, are the product of national and international collaborations. I express my deep appreciation to all scientists who were involved in these studies.

First and foremost, my gratitude goes to my supervisor, *A. Marcell Szász*, for providing me with the opportunity to conduct my research under his guidance and for consistent support throughout my PhD studies. My unofficial supervisors, *Melinda Rezeli* and *Peter Horvatovich*, thank you so much for introducing me to the world of proteomics and being my mentor throughout the years. I cannot thank you enough for your endless patience and guidance. *György Marko-Varga*, thank you for inviting me to collaborate with you and all the other excellent researchers at the European Cancer Moonshot Lund Center; your dedication to advancing melanoma research and improving patient care is truly admirable. I also extend my deepest gratitude to *David Fenyö*, whose support kick-started my PhD research in bioinformatics and for providing me countless of advice throughout the years.

I also would like to thank *Balázs Döme*, *Zsolt Megyesfalvi* and *Karin Schelch* for involving me in your SCLC studies and sharing your expertise on lung cancer. Thank you, *Lilla Turiák*, *Mirjam Balbisi* and *Gábor Tóth* for providing me the *ALK*-rearranged lung cancer dataset to work on, *Zoltán Herold* for the helpful statistical advice and *Anna Mária Tőkés* for our collaborative work on breast cancer. It has been a pleasure working with all of you.

My involvement in melanoma research at the European Cancer Moonshot Lund Center allowed me to collaborate with many excellent and dedicated researchers from all around the world. I appreciate each of you for bringing your unique perspectives into the studies. *Erika Velasquez*, you taught me so much about sample preparation in the early days; thank you for being so patient with me. Special thanks to former PhD students, *Nicole Woldmar* and *Magdalena Kuras* for our work together and for making the challenging times more bearable. *Lazaro Betancourt* and *Jeovanis Gil*, thank you for your mentorship and discussions on melanoma. *Zsolt Horvath*, your assistance with computer issues has been invaluable. *Krzysztof Pawłowski*, and *Elisabet Wieslander*, thank you for the insightful discussions on melanoma biology. *Letícia Szadai* and *István Németh*, thanks for sharing your expertise in melanoma patient care. *Yanick Hagemeijer*, *Victor Guryev*

DOI:10.14753/SE.2024.3031

and members of the Pankotai lab, *Tibor Pankotai*, *Zsuzsanna Újfaludi* and *Zoltán Páhi,* thank you for the collaborative bioinformatic work and fresh perspectives on cancer research. *Henriett Oskolás*, *Aniel Sanchez*, *Indira Pla*, *Roger Appelqvist*, *Jonatan Eriksson*, *Kim Yonghyo*, *Boram Lee* and *Johan Malm*, thank you for the pleasant chats and your assistance during my time in Lund. And lastly, my collaboration with György Marko-Varga's group at Lund University would not have been possible without the support from the *Fru Berta Kamprads Stiftelse* (Lund, Sweden).

My former supervisors and teachers at University of Pannonia, *Krisztián Horváth*, *Diána Lukács*, *Hajnalka Jankovics*, thank you for your support during my early encounters with (bio)analytical chemistry. Your enthusiasm for science made me interested in pursuing research.

Lastly, my heartfelt thanks to my family and friends for believing in me and understanding how much this academic journey means to me. I dedicate this thesis to all of you: To my *parents* and sisters*, Dóri* and *Kinga*, because I could not have done it without your unwavering support and encouragement. To my partner, *Patrik*, for being by my side during stressful times and always reminding me to take a break. To my nephews, *Ádám*, *Milán* and *Máté*, because you taught me how to enjoy the little things again. And finally, to my *pets* for their always comforting presence.