SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3491.

BALBISI MIRJAM

A gyógyszerészeti tudományok korszerű kutatási irányai

című program

Programvezető: Dr. Antal István, egyetemi tanár

Témavezetők: Dr. Turiák Lilla, tudományos főmunkatárs

MASS SPECTROMETRY-BASED CHARACTERIZATION OF PROTEOMIC AND GLYCOSYLATION ALTERATIONS IN LUNG CANCER TISSUES AND EXTRACELLULAR VESICLES

PhD thesis

Mirjam Balbisi

Semmelweis University Doctoral College

Pharmaceutical Sciences and Health Technologies Division

Supervisor:	Lilla Turiák, PharmD, Ph.D
Official reviewers:	László Márk, D.Sc Olivér Ozohanics, Ph.D
Head of the Complex Examination Committee:	Zoltán Benyó, MD, D.Sc
Members of the Complex Examination Committee:	Krisztina Ludányi, Ph.D Gitta Schlosser, Ph.D

Budapest
2026

Table of contents

List of abbreviations

AC	adenocarcinoma
ACN	acetonitrile
ALK	anaplastic lymphoma kinase, receptor tyrosine kinase
AmBic	ammonium bicarbonate
ANL	adjacent normal lung
BEGM	bronchial epithelial cell growth medium
BRAF	B-Raf proto-oncogene, serine/threonine kinase
C18	octadecyl-modified silica gel
CE	capillary electrophoresis
CID	collision-induced dissociation
CS/DS	chondroitin sulfate/dermatan sulfate
CSPG	chondroitin sulfate proteoglycan
CSPG2	versican core protein
DDA	data-dependent acquisition
DIA	data-independent acquisition
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ECM	extracellular matrix
EGFR	epidermal growth factor receptor
EML4	echinoderm microtubule-associated protein-like 4
ERGIC-53	ER-Golgi intermediate compartment 53 kDa protein
ESI	electrospray ionization
EV	extracellular vesicle
FA	formic acid
FBS	fetal bovine serum
FC	fold change
FFPE	formalin-fixed, paraffin-embedded
Fuc	fucose
FUT8	fucosyltransferase 8
GAG	glycosaminoglycan
Gal	galactose

GalNAc	*N*-acetylgalactosamine
GlcA	glucuronic acid
GlcNAc	*N*-acetylglucosamine
GSEA	gene set enrichment analysis
HA	hyaluronan
HER2	human epidermal growth factor receptor 2, receptor tyrosine kinase
Hex	hexose
HexNAc	*N*-acetylhexosamine
HFBA	heptafluorobutyric acid
HILIC	hydrophilic interaction chromatography
HPLC	high-performance liquid chromatography
HS	heparin/heparan sulfate
HSPG	heparan sulfate proteoglycan
IAA	iodoacetamide
IdoA	iduronic acid
ITA2	integrin α2
ITA3	integrin α3
ITB1	integrin β1
KRAS	Kirsten rat sarcoma viral oncogene homolog, proto-oncogene, GTPase
KS	keratan sulfate
LAMA2	laminin subunit α2
LAMA5	laminin subunit α5
LAMC1	laminin subunit γ1
LC	liquid chromatography
LCC	large cell carcinoma
LG3BP	galectin-3-binding protein
m/*z*	mass/charge ratio
MALDI	matrix-assisted laser desorption/ionization
Man	mannose
MET	MET proto-oncogene, receptor tyrosine kinase
MRPS	microfluidic resistive pulse sensing
MS	mass spectrometry

MS/MS	tandem mass spectrometry
MSI	mass spectrometry imaging
NES	normalized enrichment score
Neu5Ac	*N*-acetylneuraminic acid, sialic acid
NSCLC	non-small cell lung cancer
ofCS	oncofetal chondroitin sulfate
PASEF	parallel accumulation-serial fragmentation
PBS	phosphate buffered saline
PCA	principal component analysis
PD-1	programmed cell death protein 1
PD-L1	programmed death-ligand 1
PG	proteoglycan
poly-LacNAc	poly-*N*-acetyllactosamine
PTM	post-translational modification
Q	quadrupole
RET	rearranged during transfection, proto-oncogene, receptor tyrosine kinase
RNA	ribonucleic acid
ROS1	ROS1 proto-oncogene, receptor tyrosine kinase
SCLC	small cell lung cancer
sEV	small extracellular vesicle
SPE	solid-phase extraction
SqCC	squamous cell carcinoma
TEM	transmission electron microscopy
TFA	trifluoroacetic acid
TFPI2	tissue factor pathway inhibitor 2
THRB	prothrombin
TIMS	trapped ion mobility spectrometry
TOF	time-of-flight
UHPLC	ultra-high-performance liquid chromatography
VAR2CSA	variant surface antigen 2 chondroitin sulfate A
VEGF	vascular endothelial growth factor
WAX	weak anion exchange

1. Introduction

1.1. Mass spectrometry-based proteomics

Proteomics is the study of the proteome, which is the entire set of proteins expressed by the genome (1). In addition to the quality and quantity of proteins, the complexity of the proteome largely arises from post-translational modifications (PTMs), which regulate protein activity, conformation, localization, stability, and molecular interactions (2). PTMs are not directly encoded at the genomic level and therefore cannot be determined solely from deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) analyses. The proteome is highly dynamic and reflects the physiological state of the organism (3, 4). Therefore, comparing the proteomic profiles of healthy and tumor samples allows the identification of molecular differences that may serve as potential biomarkers, while also providing insight into the biological mechanisms of tumor development (5-7).

Proteomic workflows can be implemented using different mass spectrometry strategies: bottom-up (8), middle-down (9) or top-down (10). The most common approach is bottom-up proteomics, where proteins are enzymatically digested into smaller peptides, usually with trypsin, which cleaves at the C-terminal of lysine and arginine residues. Despite its widespread applicability, bottom-up proteomics has limitations, including the ambiguity of protein inference from shared peptides (11).

With the development of high-performance liquid chromatography (HPLC) and mass spectrometry (MS), HPLC-MS has become the primary method for proteomic separation and detection (12). These technologies require minimal sample quantities, enable rapid analysis, and allow for the simultaneous identification and quantification of large numbers of proteins. Two ionization techniques widely used for peptide and protein analysis are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI), both of which produce intact protonated or deprotonated molecules without significant fragmentation (13). In proteomics, nano-ESI is often used in combination with nano-UHPLC to increase ionization efficiency and sensitivity, especially when sample quantities are limited (14, 15). More recently, micro-flow LC-MS approaches have gained attention, offering improved reproducibility and higher throughput at the cost of a slight decrease in sensitivity (16).

Common mass analyzers include quadrupole (Q), time-of-flight (TOF), Orbitrap, and Fourier-transform ion cyclotron resonance (17). Recent developments have focused on

hybrid instrument configurations and enhanced analyzer designs, such as systems integrating ion mobility separation (e.g., timsTOF) or combining multiple high-speed mass analyzers (e.g., Orbitrap Astral platforms) (18, 19).

Tandem mass spectrometry (MS/MS) plays a crucial role in the fragmentation of molecules, enabling the determination of peptide sequences and thus the identification of proteins (20). Fragmentation can be induced by collisions with gas molecules (collision-induced dissociation, CID), electrons, or photons, with CID being the most widely used method in bottom-up workflows (21). Data collection can be data-dependent (DDA) or data-independent (DIA) (22, 23). In DDA, the software selects the most intense peaks for fragmentation and generates MS/MS spectra only for these peptides. Although widely used, DDA does not identify low-abundance peptides and has relatively low reproducibility. In DIA, the mass range is divided into consecutive windows and all ions within each window are fragmented, providing more comprehensive and reproducible peptide coverage, however, it requires more complex data analysis.

Peptides and proteins are typically identified by database searching, where theoretical spectra are generated for all possible peptides obtained by the theoretical digestion of all proteins in the database and compared to experimental data (24). Each peptide spectrum pair is assigned a score and peptides above a threshold are accepted as identified. Protein identification is then inferred from the collection of identified peptides. Quantification can be performed using labeling or label-free methods (25, 26). Label-free quantification based on MS/MS peak intensities is simple, cost-effective, and widely used, although it only allows for relative quantification. In contrast, labeling workflows enable absolute protein quantification but are associated with higher cost and more complex experimental design.

1.2. Protein glycosylation

Protein glycosylation is one of the most structurally complex PTMs in mammalian biology, occurring in more than 50% of proteins (27). It has a major impact on protein folding, stability, cell-cell recognition, receptor activation, immunomodulation, and intracellular signaling (27, 28). Unlike DNA and proteins, whose sequences are template-driven, glycans are synthesized through a non-template-based enzymatic network, resulting in enormous structural heterogeneity (29). Major glycosylation types

include *N*-glycosylation, *O*-glycosylation, and glycosaminoglycans (GAGs), of which *N*-glycosylation and GAGs are discussed in detail below.

There is growing evidence that changes in protein glycosylation influence cancer biology at multiple, interconnected levels (30-34). In early tumor development, dysregulated glycosylation alters signaling pathways that regulate cell proliferation by modifying key regulatory proteins such as EGFR, thereby contributing to uncontrolled cell growth. Altered glycosylation patterns of adhesion-related proteins (e.g., integrins, cadherins) weaken normal cell-cell and cell-matrix interactions, allowing malignant cells to break through these boundaries. Altered glycosylation of angiogenic factors, including VEGF, contributes to enhanced angiogenic capacity and improved metabolic support for growing tumors. In parallel, cancer cells actively modify the glycosylation of immune checkpoint proteins, reducing immune detection and promoting immune escape. In advanced stages, changes in the glycosylation of proteins involved in motility, adhesion, and extracellular matrix (ECM) remodeling promote cell spread and invasion, ultimately supporting metastatic spread and colonization of distant sites.

The clinical relevance of protein glycosylation is further underscored by the fact that many widely used cancer biomarkers are glycoproteins, including α-fetoprotein, prostate-specific antigen, human epidermal growth factor receptor 2, and carcinoembryonic antigen, and in several cases, their glycoforms provide additional diagnostic value (35-37).

1.2.1. *N*-glycosylation

In *N*-glycosylation, *N*-glycan chains are covalently linked to the amide nitrogen of asparagine residues within the asparagine-X-serine/threonine consensus sequence, where X can be any amino acid except proline (38). In human proteins, all *N*-glycans contain a common pentasaccharide core consisting of two *N*-acetylglucosamine (GlcNAc) and three mannose (Man) residues. This core structure can be further extended, forming three main classes of *N*-glycans: oligomannose, hybrid, and complex types. Oligomannose *N*-glycans contain only Man residues outside the core, while complex *N*-glycans have more elaborate antennae containing GlcNAc, galactose (Gal), and often terminal sialic acid (*N*-acetylneuraminic acid, Neu5Ac) residues. In addition, fucose (Fuc) can be attached to the core GlcNAc or to the antenna GlcNAc units, further increasing structural diversity. More complex structures can also form, including multi-antennary and highly

branched structures, as well as elongated poly-*N*-acetyllactosamine (poly-LacNAc) chains. A schematic representation of *N*-glycan structures is shown in **Fig. 1**.

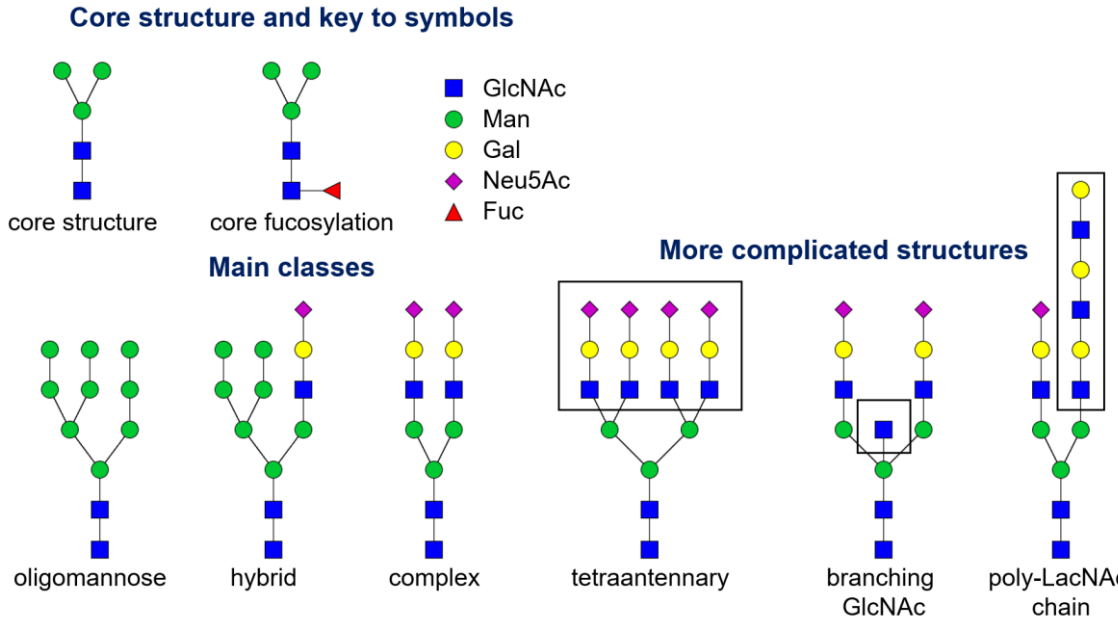

Figure 1. Structure of *N*-glycans: core structure, main classes, and examples of more complicated structures.

N-glycan biosynthesis begins in the endoplasmic reticulum with the assembly of a lipid-bound oligosaccharide precursor, which is then transferred to the nascent polypeptides. This is followed by sequential trimming and processing steps in the endoplasmic reticulum and Golgi apparatus, where glycosidases and glycosyltransferases generate a wide variety of oligomannose, hybrid, and complex *N*-glycan structures.

Two complementary strategies are commonly used to study *N*-glycosylation (39). In *N*-glycomics, *N*-glycans are enzymatically released from proteins, allowing global profiling of *N*-glycan composition independently of their protein backbone. In contrast, *N*-glycoproteomics is based on proteolytic digestion followed by the enrichment and analysis of intact *N*-glycopeptides, which provides site-specific information on *N*-glycan occupancy and microheterogeneity. Although site-specific analysis provides greater biological insight, released *N*-glycan profiling remains more widely used as *N*-glycopeptide analysis presents significant challenges, including the need for efficient enrichment and data evaluation strategies. In recent years, significant progress has been made in the analysis of intact *N*-glycopeptides, thanks to advances in enrichment strategies (40, 41), fragmentation (42, 43) and data acquisition methods (44, 45), and

bioinformatics tools (46). As a third approach, intact *N*-glycoproteins can also be analyzed, however, its application is limited in complex samples due to the large variety of structures (47).

1.2.2. Glycosaminoglycans

Proteoglycans (PGs) are abundant components of the ECM and the cell surface, where they contribute to cell signaling, tissue structure, and the regulation of inflammatory processes (48). Their defining feature is the presence of one or more GAG chains covalently attached to a protein backbone. GAGs are linear polysaccharides composed of repeating disaccharide units (hexuronic acid/Gal + *N*-acetylhexosamine, HexNAc) that can be sulfated at distinct positions (49). Based on the composition of their repeating units, GAGs are classified into several families, including hyaluronan (HA), keratan sulfate (KS), chondroitin sulfate/dermatan sulfate (CS/DS), and heparin/heparan sulfate (HS) (50).

Among these, CS/DS and HS are the most extensively studied due to their prominent roles in cell signaling and disease processes. CS is composed of repeating disaccharides of glucuronic acid (GlcA) and *N*-acetylgalactosamine (GalNAc), while DS contains iduronic acid (IdoA) formed through epimerization of GlcA. The CS/DS GalNAc residues can be *O*-sulfated at the C4 and C6 positions. HS chains consist of repeating disaccharide units of GlcA or IdoA and GlcNAc. Sulfation can occur at the C2 position of the uronic acid, the C6 position of the GlcNAc, and at the amino group, where *N*-acetyl groups can be replaced by *N*-sulfate groups. The possible structures and their Lawrence code abbreviations are shown in **Fig. 2** (51).

GAG biosynthesis occurs in the Golgi apparatus through the stepwise elongation of polysaccharide chains on specific serine residues of the core protein (48). This process is initiated by the formation of a conserved tetrasaccharide linker region (GlcA-Gal-Gal-xylose-*O*-serine), followed by the alternating addition of monosaccharide units carried out by specific glycosyltransferases. Subsequent modifications, including the epimerization of GlcA to IdoA and sulfation at specific positions, are catalyzed by epimerases and sulfotransferase enzymes.

The analytical characterization of GAGs is most often performed using bottom-up strategies, in which polysaccharide chains are digested chemically or by enzymes. Lyases cleave via a β-elimination mechanism, resulting in the formation of an unsaturated group

at the non-reducing end of the molecule, while the use of hydrolase enzymes results in the formation of saturated disaccharides (52). The HPLC-MS characterization of GAGs is hindered by their strong negative charge, labile sulfate groups, and wide polarity range (52, 53). In our research group, we use in-house developed, hydrophilic interaction chromatography and weak anion exchange (HILIC-WAX) based chromatography methods for the separation of CS/DS and HS disaccharides (54, 55).

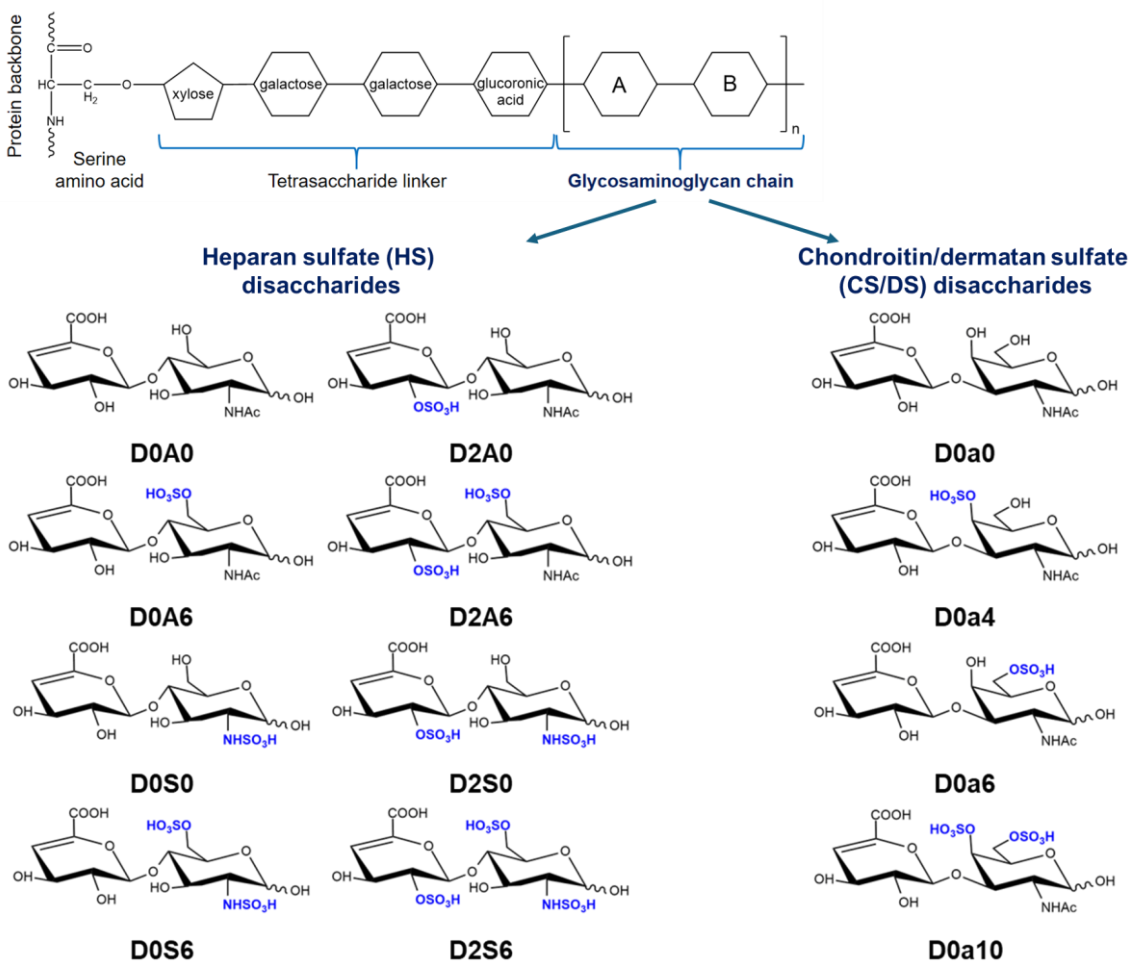

Figure 2. Structure of proteoglycans, along with the structure and Lawrence code designation of disaccharides produced during the digestion of HS and CS/DS chains by lyase enzymes.

1.3. Lung cancer

In 2022, lung cancer was the most frequently diagnosed cancer worldwide and the leading cause of cancer-related mortality, with approximately 2.5 million new cases and 1.8 million deaths (56). Lung cancer is classified as small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), with SCLC accounting for approximately 11% of

diagnosed cases and NSCLC accounting for 89% (57). These two types differ in their clinical management: SCLC is typically not amenable to surgical resection but shows initial sensitivity to chemotherapy and radiotherapy, whereas NSCLC may be treated surgically in early stages (58, 59).

NSCLC represents a heterogeneous group of diseases, comprising several histological subtypes, most notably adenocarcinoma (AC, 51%), squamous cell carcinoma (SqCC, 25%), and large cell carcinoma (LCC, 6%) (57). Among these, AC is characterized by a high prevalence of oncogenic driver mutations, including mutations or rearrangements on the KRAS, EGFR, MET, HER2, BRAF, RET, ROS1, and ALK genes (60).

Anaplastic lymphoma kinase (ALK) is a receptor tyrosine kinase involved in cell signaling, with important roles in nervous system development (61). Around 4% of NSCLC patients carry a rearrangement in the ALK gene on chromosome 2, resulting in the formation of oncogenic fusion proteins, e.g., EML4-ALK, causing abnormal activation of downstream signaling cascades (62). The detection of genetic alterations is routinely performed using molecular pathology techniques. Fluorescence in situ hybridization and immunohistochemistry are gold standards for detecting ALK translocation (63). The frequency of lung cancer types and major gene mutations is illustrated in **Fig. 3**.

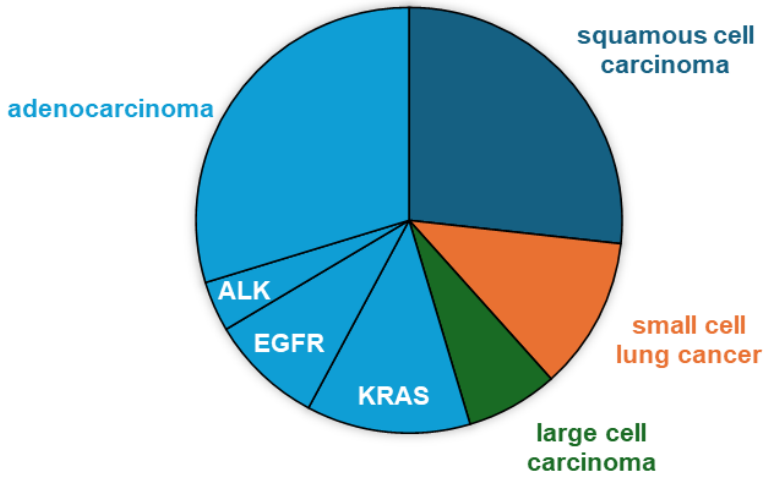

Figure 3. The incidence of different types of lung cancer and the three most common gene rearrangements underlying adenocarcinoma.

With the development of precision oncology, molecularly targeted therapies and immunotherapies have become part of standard clinical practice (60, 64). Patients whose

tumors have treatable genetic alterations can benefit from targeted inhibitors that target specific oncogenic factors, such as tyrosine kinase inhibitors against EGFR mutations or ALK rearrangements (65, 66). In parallel, immune-based therapies – particularly immune checkpoint inhibitors targeting the PD-1/PD-L1 interaction – have shown significant clinical benefits in several lung cancer subtypes (67). Despite these therapeutic advances, the emergence of acquired drug resistance to both targeted and immune-based therapies remains a significant challenge in the long-term treatment of lung cancer patients (68, 69). This therapeutic limitation contributes to lung cancer mortality and highlights the need for ongoing research into the molecular and cellular mechanisms underlying tumor progression, treatment response, and resistance.

1.4. Glycosylation in lung cancer

There is growing experimental evidence that glycosylation is significantly altered in lung cancer and contributes to the development of tumor-specific molecular characteristics. Changes in *N*-glycan composition and site-specific *N*-glycosylation were reported in several tissue-based studies, typically after tissue homogenization and digestion in-solution. For example, integrated proteomic, *N*-glycoproteomic, and *N*-glycomic analyses of lung AC tissues revealed significant differences between tumor and non-tumor samples, including an increase in the relative abundance of oligomannose, fucosylated, and sialofucosylated *N*-glycan structures in the tumor and changes in the *N*-glycosylation of proteins involved in metabolism, cell adhesion, and regulatory pathways (70).

In addition to tissues, numerous studies examined *N*-glycosylation changes in blood samples (primarily blood serum), e.g., *N*-glycomic profiling of lung cancer and non-cancerous patients identified increased amounts of different tri- and tetraantennary *N*-glycans in lung cancer (71). In another study, differential representation of *N*-glycopeptides derived from proteins involved in complement activation, coagulation, and inflammatory responses was found (72). Although tissue and blood remain the primary sample types, alternative biological fluids such as saliva (73), pleural effusion (74), and bronchoalveolar lavage (75) also showed promise as sources of lung cancer-associated *N*-glycosylated biomarkers.

Compared to *N*-glycosylation, significantly fewer studies addressed GAGs in lung cancer. GAG disaccharide analysis of tissues from SqCC patients showed that tumor tissue

samples contained more than twice as much CS/DS as normal tissues and had an altered sulfation pattern, while no significant change was observed for HS and HA (76). In a GAG study conducted by our research group, a general increase in CS/DS amount was observed in tumor tissues, accompanied by changes in sulfation patterns, while different histological subtypes differed in the 6S/4S sulfation ratios (77). Small differences were observed between lung AC samples with different oncogenic mutations, but CS/DS patterns were not grouped according to mutation type (78).

A comprehensive overview of these findings and the associated analytical strategies is provided in our review, highlighting both existing knowledge and gaps that remain in the field (34).

1.5. Spatial characterization of tumor tissues

Tumors are highly heterogeneous structures composed of malignant cells together with diverse non-malignant components, including immune cells and stromal cells, each of which may exhibit distinct molecular and glycosylation profiles (79). In lung cancer, tumors often show marked histological variability, and different morphological patterns such as tubular, papillary, or solid tumor structures may coexist within the same tumor. Papillary tumors are characterized by cuboidal tumor cells arranged along fibrovascular cores, forming finger-like projections, while tubular tumors show gland-forming structures, in which malignant cells organize into gland-like or tubular architectures within a fibrous stroma (80). In contrast, solid tumors lack glandular differentiation and consist of densely packed tumor cells. These patterns have been associated with differences in tumor progression and patient prognosis (81, 82). In addition, tumor regions can vary in terms of stromal content and mucin production. Stromal regions contain ECM components, e.g., PGs that influence tissue mechanics and cell signaling, while mucin-rich regions are characterized by the presence of heavily *O*-glycosylated mucin proteins, which are frequently overexpressed in cancers (83, 84).

Workflows based on tissue homogenization provide valuable information about global changes but miss information about the spatial heterogeneity of complex tumor microenvironments. As a result, analytical strategies that preserve spatial information are becoming increasingly important for a more comprehensive understanding of tumors.

Traditional spatial imaging methods used in pathology include immunohistochemistry and immunofluorescence, which enable the visualization of specific proteins and cellular

markers within tissue sections (85). These antibody-based techniques have become essential tools in both research and diagnostics and form the basis of large-scale databases such as the Human Protein Atlas, which systematically maps protein expression patterns across normal and cancer tissues (86). However, antibody-based approaches are limited to predefined targets and generally provide restricted molecular coverage.

More recently, MS-based spatial omics technologies have emerged that allow broader molecular profiling while preserving tissue architecture. MALDI-mass spectrometry imaging (MSI) is highly effective for the spatial mapping of small molecules, including lipids, metabolites, and released glycans (87). Protein analysis generally relies on on-tissue digestion workflows followed by peptide-level MSI, while the analysis of intact proteins remains considerably more challenging due to their larger size, lower ionization efficiency, and the complexity of tissue samples. Emerging approaches, such as Deep Visual Proteomics, which combines high-resolution imaging, artificial intelligence-guided single-cell isolation, laser microdissection, and ultrasensitive LC-MS/MS proteomics, have further expanded the possibilities for spatially resolved proteomic analysis at near single-cell resolution (88).

MSI of released *N*-glycans was successfully applied to lung AC tissues, revealing specific *N*-glycans enriched in areas associated with immune cell infiltration, fibrosis, or necrosis (89). Despite significant methodological advances, the MSI analysis of intact glycopeptides is still impossible due to limited sensitivity and the lack of effective enrichment strategies. Although recent proof-of-concept studies have demonstrated that spatially resolved detection of GAG disaccharides is feasible using advanced MALDI-based ionization strategies, these methods are still technically demanding and have limited applications (90, 91).

To overcome these limitations, our research group developed surface digestion workflows in which enzymatic digestion occurs directly on tissue sections (92). Although this approach does not provide pixel-level spatial resolution like MSI, it allows for the targeted analysis of multiple spatially defined tissue regions and the investigation of analyte classes that are currently not accessible with MSI (93).

1.6. Extracellular vesicles

Extracellular vesicles (EVs) are membrane-enclosed particles that are actively secreted by cells into the extracellular space and represent an important mechanism of intercellular

communication (94, 95). EVs carry a diverse molecular cargo, including lipids, proteins, and nucleic acids, thereby mediating the transfer of biological information between cells. Through this function, EVs play a key role in the regulation of both physiological and pathological processes, including cancer initiation, progression, and metastasis (96). According to the current recommendations, EVs are classified according to their size into small EVs (sEVs, <200 nm) and large EVs (>200 nm), rather than based on their biogenesis (97).

Small EVs are of particular interest in oncology, as they reflect the molecular characteristics of their parent cells and are actively involved in tumor-related processes such as angiogenesis, immune modulation, invasion, metastasis, and therapy resistance (98). They can be detected in a variety of body fluids, including blood, urine, and saliva, making them attractive for the discovery of minimally invasive tumor biomarkers (99). These EV-based biomarker assays primarily focus on nucleic acids and proteins. Proteomic profiling of EVs provided important insights into lung cancer-associated molecular signatures. For example, plasma EVs of NSCLC patients were enriched in proteins linked to pathways related to actin cytoskeleton regulation, focal adhesion, and ECM-receptor interactions (100). Differences in EV protein composition were observed between major histological subtypes of lung cancer, indicating that EV cargo may also reflect tumor heterogeneity (101).

In contrast, PTMs of EV proteins remain largely uncharacterized (102). This is primarily due to technical challenges related to the limited amount of starting material and the analytical complexity of modified proteins. However, given the central role of glycosylation in protein function and cancer biology, the systematic characterization of EV-associated glycoproteins is a promising area for both the discovery of biomarkers and the advancement of our understanding of tumor biology. In line with this, previous work showed that the *N*-glycosylation profiles of sEVs derived from SCLC and NSCLC cells differ substantially (103).

2. Objectives

The aim of this dissertation was to investigate changes in protein expression and glycosylation in biological systems related to lung cancer using HPLC-MS-based analytical workflows, with a particular focus on the characterization of GAG disaccharides.

The dissertation is based on two related projects. The first project focused on investigating proteomic profiles and GAG compositions (including CS/DS and HS) in different regions of formalin-fixed, paraffin-embedded (FFPE) lung tissue sections carrying ALK rearrangement (104). The tissue regions were characterized by a pathologist based on their morphology, mucin and stromal content. The aim was to identify and quantify proteins and GAG disaccharides in different regions, and to reveal alterations between regions with different tumor morphology and tumor microenvironment characteristics that may contribute to a better understanding of lung cancer biology. The analysis was performed using on-surface tissue digestions followed by solid-phase extraction (SPE) purifications and HPLC-MS measurements. The resulting datasets were evaluated using statistical and bioinformatic approaches, and the findings were interpreted in their biological context.

Although tissue samples are valuable for studying disease-related molecular changes, their clinical diagnostic application is limited due to the invasive nature of tissue sampling. Therefore, the second project focused on EVs, which are promising sources of biomarkers because they can be isolated from minimally invasive biological samples such as blood. Previous studies have shown that the proteomic composition of EVs changes significantly in cancer; however, their glycosylation patterns remain poorly characterized, partly due to the limited sample amounts typically available. In particular, GAGs have been rarely studied in EVs.

To address this gap, sEVs derived from A549 lung AC and BEAS-2B non-tumorigenic epithelial cells were used as a model system (105). Our goal was to isolate and characterize sEVs, and to confirm their size distribution and morphology using microfluidic resistive pulse sensing (MRPS) and transmission electron microscopy (TEM). Subsequently, comprehensive multi-omics characterization of the sEV samples was performed, including proteomics, *N*-glycoproteomics, and CS/DS disaccharide analysis using HPLC-MS. Finally, statistical and bioinformatic methods were applied to

identify significantly altered proteins, *N*-glycopeptides, and GAG characteristics, and to interpret these alterations in a biological context.

An overview of the work process is shown in **Fig. 4**. Please note that the figure is schematic and not all analyses were performed for each sample type: HS disaccharides were analyzed only from tissues, while *N*-glycoproteomics was performed only for EV samples.

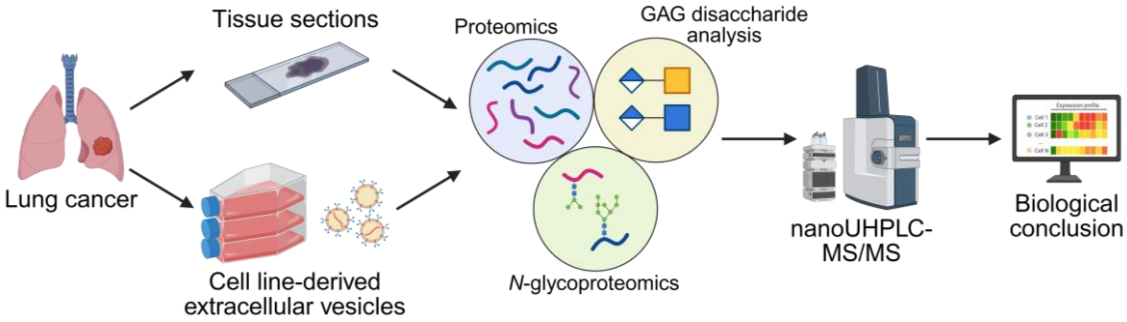

Figure 4. Schematic representation of the experimental workflows used in this dissertation. Lung tissue sections and EVs derived from bronchial epithelial and lung cancer cell lines were analyzed using HPLC-MS-based multi-omics (proteomics, GAG disaccharide analysis, *N*-glycoproteomics), followed by bioinformatic evaluation to identify biological alterations. Created with BioRender.com.

3. Methods

3.1. Tissue study

3.1.1. Tissue samples and preparation

In the first study, proteomic as well as CS/DS and HS disaccharide analyses were performed on 22 tissue regions from 7 FFPE tissue sections. The study was approved by the Institutional Review Board of the Hungarian Ministry of Health (TUKEB numbers: 2521-0/2010-1018EKU, 52614-4-213EKU). The tumor regions were characterized by pathologists based on morphological classification, mucin and stromal content, and the characteristics of the regions are summarized in **Table 1**.

Table 1. Morphology, mucin and stromal content of lung tissue regions. Mucin or stroma scores of 1, 2, and 3 represent mucin or neostroma covering 1-33%, 33-66%, and 66-100% of the total area, respectively.

Patient number	Region number	Morphological classification	Mucin (1-3)	Stroma (1-3)	Patient number	Region number	Morphological classification	Mucin (1-3)	Stroma (1-3)
1	1	Tubular	3	3	4	12	Papillary	3	2
1	2	Tubular	3	3	5	13	Solid	–	2
2	3	Solid	2	2	5	14	Solid	–	2
2	4	Normal	–	–	5	15	Solid	–	3
3	5	Normal	–	–	5	16	Solid	–	–
3	6	Normal	–	–	6	17	Solid	–	1
3	7	Tubular	2	1	6	18	Solid	–	1
3	8	Papillary	2	1	6	19	Solid	–	1
3	9	Papillary	2	3	7	20	Solid	1	3
4	10	Papillary	3	2	7	21	Solid	–	3
4	11	Papillary	3	3	7	22	Normal	–	–

6 µm thick sections were mounted on SuperFrost Ultra Plus microscope slides. The tissue sections were heated at 60 °C for 2 h to promote adhesion, then washed sequentially (xylene 2 × 5 min, ethanol 2 × 3 min, 90% ethanol 3 min, 70% ethanol 3 min, 10 mM ammonium bicarbonate (AmBic) 5 min, and water 1 min) to remove the paraffin and rehydrate the tissues. The formalin induced cross-links were then broken by heating the

tissue sections in citrate buffer (94.6 mM sodium citrate, 20.8 mM citric acid, pH = 6) at 80-85 °C for 30 min. The borders of the regions of interest were marked with a razor blade on the sections and then tissue surface digestion was performed. The proteomic and GAG digestions were performed on separate, parallel slides.

3.1.2. Proteomic analysis

3.1.2.1. On-tissue proteolytic digestion and peptide purification

For tryptic digestion, the S-S cross-links were reduced first by pipetting 2 µL 0.1% RapiGest + 5 mM dithiotreitol (DTT) + 20% glycerol and incubating at 55 °C for 20 min in a humidified chamber. Next, alkylation of the cysteine residues was performed by 2 µL 25 mM AmBic + 10 mM iodoacetamide (IAA) + 20% glycerol and kept in the dark for 20 min at room temperature. Digestion was performed in 5 consecutive 40 min cycles at 37 °C (92). In the first two cycles, 2 µL 40 ng/µL Lys-C/trypsin in 50 mM AmBic and 15% glycerol was applied, while the following three cycles consisted of 2 µL 200 ng/µL trypsin in 50 mM AmBic and 15% glycerol. At the end of digestion, the liquid was pipetted from the tissue regions, and then the remaining peptides were extracted from the surface by pipetting up and down with 2 µL 10% acetic acid five times. Finally, the solvents were evaporated and stored at −20 °C until purification.

Samples were purified on C_{18} spin cartridges, with centrifugation at 2000 rpm for 1 min at each step (106, 107). Solvents were held at 4 °C during the purification, except the elution solvents which were kept at room temperature. First, the cartridges were equilibrated with 2 × 200 µL 50% methanol, 2 × 200 µL 0.5% trifluoroacetic acid (TFA) + 5% acetonitrile (ACN), 2 × 200 µL 0.1% heptafluorobutyric acid (HFBA). Next, samples were applied in 200 µL 0.1% HFBA and the flow-throughs were reapplied. Cartridges were washed with 2 × 100 µL 0.1% HFBA, followed by elution of the purified peptides with 2 × 50 µL 0.1% TFA + 70% ACN and 50 µL 0.1% formic acid (FA) + 70% ACN. Samples were dried down and stored at −20 °C.

3.1.2.2. LC-MS/MS analysis of peptides

Samples were measured using a Maxis II Q-TOF (Bruker, Germany) MS system connected to a Dionex Ultimate 3000 RSLC nano-UHPLC (Thermo Fischer Scientific, USA). Purified peptides were dissolved in 8 µL 0.1% FA + 2% ACN and 6 µL was injected. Peptides were trapped on a Thermo Fischer Acclaim PepMap100 C_{18} column (5 µm, 100 µm x 20 mm) and separated on a Waters Acquity M-Class BEH130 C_{18}

analytical column (1.7 µm, 75 µm × 250 mm) at 48 °C with a solvent system of 0.1% FA in water (A) and 0.1% FA in ACN (B). Separation was performed at a flow rate of 0.3 µL/min, starting with 4% B, which was increased to 25% over 75 min, 40% over 15 min, and 90% over 1 min, held for 5 min, then equilibrated with 4% B for 20 min.

A CaptiveSpray nanoBooster ion source was used in positive mode, with 0.2 bar gas pressure, 3 L/min drying gas flow rate and 1150 V capillary voltage. MS and MS/MS spectra were collected using DDA with a cycle time of 2.5 s in the 150-2200 *m*/*z* range, an MS1 data acquisition frequency of 3 Hz and an MS/MS data acquisition frequency of 4 or 16 Hz depending on the precursor intensity. Based on the precursor *m*/*z* value and charge, the collision energy was calculated by the software. The mass values were recalibrated using Compass DataAnalysis software, with sodium formate injected in parallel with the sample.

3.1.2.3. Proteomic data processing and statistical analysis

Peptides and proteins were identified using Byonic (108) based on the Swiss-Prot human protein database (accessed November 2020). A precursor ion mass tolerance of 5 ppm and a fragment ion mass tolerance of 10 ppm were used, with a maximum of 2 missed cleavages. Carbamidomethylation of cysteine was a fixed modification, while deamidation of asparagine and glutamine, and oxidation of methionine were variable modifications. Protein identifications were accepted if at least two unique peptides were detected with a LogProb greater than 1.3 and the protein-level LogProb exceeded 2. The proteins were then quantified using MaxQuant (109) with the protein database identified by Byonic, using the same modifications, maximum 2 missed cleavages, a first search with 20 ppm and a main search with 10 ppm, LFQ algorithm and the match between runs function.

Statistical analysis and data visualization were performed using custom code in R, assisted by RStudio. Graphs were created using the ggplot2 and gplots packages. Only proteins quantified in at least three samples in at least two sample groups were considered, without any imputation step. Differences in protein expression between groups were evaluated using the Wilcoxon signed-rank test. Nominal *p*-values below 0.05 were accepted as statistically significant, except for the comparison between tubular tumors and adjacent normal regions, where the lowest possible value (0.057) was used as the threshold. Clustered heatmap (heatmap.2 function, ward.D2 clustering) and principal

component analyses (PCA, prcomp function) were performed to visualize global proteomic changes. The STRING webserver was used to explore altered biological processes based on the differentially expressed proteins (110). The minimum interaction score was set to high confidence (0.700) and unconnected nodes were excluded from the network.

3.1.3. Glycosaminoglycan analysis

3.1.3.1. On-tissue GAG digestion and GAG disaccharide purification

CS/DS and HS digestions were performed sequentially. First, CS/DS chains were digested by pipetting 2 µL drops of 2 mU/µL chondroitinase-ABC + 20 mM Tris (pH = 7.6) + 2.5 mM ammonium acetate + 10% glycerol onto each region and incubating at 37 °C. This step was repeated hourly for 4 h (5 drops in total) and the last drop was left on the tissue for an additional 20 h. CS/DS disaccharides were collected by 5 × 2 µL 1% NH_3 extraction and dried down, while the tissues were washed with 10 mM Tris (pH = 7.6). Next, HS digestion was performed on the same regions by adding 5 x 2 µL 0.5 mU/µL heparin lyase I + 0.1 mU/µL heparin lyase II + 0.1 mU/µL heparin lyase III + 20 mM Tris (pH = 7.6) + 2.5 mM $Ca(OH)_2$ + 10% glycerol. Digestion was performed at 37 °C for 48 h and enzyme drops were added after 0, 1, 2, 19, and 26 h (92, 111). Finally, HS disaccharides were collected in the same way as CS/DS.

GAG disaccharides were purified using a two-step SPE workflow (112). All centrifugation steps were performed at 2500 rpm for 1 min. In the first step, self-packed cotton-HILIC pipette tips were used, which were equilibrated with 50 µL 60% ACN and 2 x 50 µL 1% TFA + 95% ACN. Samples were loaded in 30 µL 1% TFA + 95% ACN, and the flow-through was reapplied twice. The tips were washed with 50 µL 1% TFA + 95% ACN. Flow-through fractions from sample loading and washing were collected and dried. Disaccharides retained on the cotton-HILIC tips were eluted using 2 x 10 µL NH_3 solution pre-heated to 40 °C. For CS/DS disaccharides, 1% NH_3 was applied, while 5% NH_3 was used for HS disaccharides. Next, the dried flow-through fractions from the cotton-HILIC step were subjected to porous graphitic carbon-based SPE. The cartridges were washed with 2 x 100 µL 0.1% TFA + 80% ACN and 2 x 100 µL water. Samples were applied in 50 µL water, incubated for 2 min, and centrifuged, followed by reapplication of the flow-through. After washing with 3 x 100 µL water, bound analytes were eluted using 3 x 50 µL 0.05% TFA + 40% ACN.

Elution fractions obtained from both purification steps were combined, dried down, and samples were kept at −20 °C until further analysis.

3.1.3.2. LC-MS/MS analysis of GAG disaccharides

GAG samples were measured on a Waters Select Series cyclic ion mobility MS (Milford, USA) connected to a Waters Acquity I-class UHPLC. GAG disaccharides were dissolved in 8 µL 10 mM ammonium formate + 75% ACN (pH = 4.4) solution and 1.5 µL was injected. Separation was performed on a self-packed GlycanPac AXH-1 HILIC-WAX capillary column (250 µm × 10 cm) with a flow rate of 10 µL/min and temperature of 45 °C. Solvents A and B were 10 mM and 65 mM ammonium formate + 75% ACN (pH = 4.4) solution, respectively (54, 55). Gradient applied for the separation of HS disaccharides started from 6% B, increased to 100% in 0.5 min, held for 7.5 min, then decreased to 6% B in 0.5 min, and equilibrated for 16.5 min. CS/DS disaccharides were eluted with 8% B for 10 min.

The MS ion source was a low-flow ESI source used in negative mode, with source temperature of 120 °C, capillary voltage of 1.9 kV and cone voltage of 20 eV. MS1 spectra were collected in the 180-680 and 200-600 *m*/*z* ranges for HS and CS/DS, respectively. In the case of CS/DS, MS/MS spectra were also recorded to determine the ratio of the D0a4 and D0a6 isomers, with ions fragmented at a collision energy of 32 eV.

3.1.3.3. GAG data processing

Chromatographic peak integration was performed using the TargetLynx plug-in of the MassLynx control software, within a mass window of 0.08 Da. Quantities were determined based on a calibration series, and a calibration function was used to determine the ratio of the D0a4 and D0a6 isomers based on the ratio of characteristic fragment ion peaks at 282 and 300 *m*/*z*. Quantitative values were calculated from the peak areas using a calibration method, while the calibration function recorded for the known ratios was used to determine the ratio of the D0a4 and D0a6 isomer pairs. PCA and clustered heatmap analyses were performed in R, analogous to proteomics. All other figures were prepared in OriginPro.

3.2. Extracellular vesicle study

3.2.1. Cell culturing, isolation and lysis of small extracellular vesicles

A549 cells were cultured in F12 medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin, while BEAS-2B cells were cultured in bronchial epithelial cell growth medium (BEGM). BEAS-2B culture flasks were pretreated with 0.01% poly-L-lysine to promote adhesion. Cells were stored at 37 °C in a humidified incubator containing 5% CO_2. The medium was replaced with 10 mL of FBS-free medium in each T75 flask to reduce bovine-derived contaminants.

The supernatants were collected after 72 h, and larger particles (cell debris, apoptotic bodies, microvesicles) were removed by centrifugation at 10000 g for 30 min at 4 °C and subsequent filtration of the supernatant through a 0.22 μm filter. Solutions were concentrated on 10 kDa Amicon Ultra-15 filters and the resulting 1 mL solutions were applied on in-house prepared size exclusion chromatography columns packed with 10 mL Sepharose CL-2B (113). Fractions were eluted with 1-1 mL phosphate buffered saline (PBS), and fractions 3, 4, and 5 were combined for further steps. Samples were then taken to confirm the presence of sEVs in the sample based on size and structure, in accordance with Minimal information for studies of extracellular vesicles 2023 specifications (97).

Next, solvent exchange was performed on 10 kDa Amicon Ultra-0.5 centrifuge filters. Filters were washed with 200 μL water, then samples were applied in 500 μL fractions and washed with 200 μL 200 mM AmBic, followed by 200 μL 50 mM AmBic. Each step was followed by centrifugation at 13500 g for 10 min, except for the last step, which was performed for 15 min.

After that, the filters were inverted and centrifuged at 1000 g for 1 min to recover the sEV fractions in MS-compatible AmBic solution. Repeating freeze-thaw cycles were used for lysis, during which the samples were frozen in liquid nitrogen for 30 sec, followed by 10 min of ultrasonication. A total of 7 cycles were applied, replacing liquid nitrogen with 1 h of freezer in the 2^{nd} and 4^{th} cycles. Protein concentrations were determined using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA) by measuring absorbance at 280 nm. Where necessary, samples were pooled to obtain enough material (at least 15 μg/sample) for further analysis.

To account for potential background signals from the cell culture medium, control samples of F12 and BEGM media were analyzed. Additional sample preparation steps

were performed using six sEV samples from each cell line and three control samples from each culture medium. The exception was *N*-glycomic analysis, in which one representative sample from each group was analyzed.

3.2.2. Characterization of small extracellular vesicles

The size distribution was determined using MRPS technique, for which the sEV sample was diluted with pre-filtered 0.3% Tween-20-containing PBS, then loaded into C400 cartridges, and measured with an nCS1 instrument (Spectradyne LLC, USA) in the 65-400 nm measurement range.

TEM was used to characterize the sEV morphology (114). For this purpose, 3 µL samples were placed on formvar-coated grids, dried for 10 min, fixed with 2% glutaraldehyde in PBS for 10 min, and then washed with water three times, each for 5 min. Then the sEVs were contrasted with 2% methyl cellulose containing 0.4% UranyLess on ice for 10 min. Measurements were carried out using a JEM1010 (Jeol, Japan) TEM and ImageJ was used to analyze the images.

3.2.3. *N*-glycomic analysis

N-glycomic profiling was performed using a previously established capillary electrophoresis (CE)-MS workflow (115-118). Briefly, *N*-glycans were enzymatically released using a polyvinylidene fluoride membrane-based immobilization approach followed by overnight incubation with PNGase F. Released *N*-glycans were subjected to linkage-specific Neu5Ac derivatization to stabilize and differentiate Neu5Ac linkages. After derivatization, *N*-glycans were purified using cotton-HILIC SPE. *N*-glycans were subsequently labeled with Girard's reagent P, introducing a permanent positive charge. CE-MS analysis was performed in positive mode on a CESI 8000 Plus system (Sciex, USA) coupled to a Maxis Plus Q-TOF (Bruker Daltonics, Germany) via a sheathless CE-ESI interface. Separations were carried out in a bare fused-silica capillary using 20% acetic acid as background electrolyte.

N-glycan compositions were identified based on MS1 using GlycoGenius (119). A broad theoretical library was created consisting of *N*-glycans with 3-10 hexose (Hex), 0-2 Fuc, 2-8 HexNAc, 0-4 Neu5Ac, and 0-1 phosphorylation/sulfation. Girard P reducing-end tagging and Neu5Ac derivatization were included in the search parameters, and protonated species with 1-3 charges were considered. Candidate compositions were filtered for minimum isotopic fitting score of 0.8, minimum curve fitting score of 0.9,

minimum signal-to-noise ratio of 3, and a mass error within −10 to +15 ppm. The resulting annotations were further refined based on electrophoretic migration behavior, and the resulting list of *N*-glycan compositions was then used to create a database for *N*-glycopeptide search.

3.2.4. Proteomic and *N*-glycoproteomic analysis

3.2.4.1. Digestion, *N*-glycopeptide enrichment and peptide purification

For proteomic analysis, 5 µg protein from each sample was diluted with water to 35 µL. Then, 1.5 µL methanol, 2 µL 200 mM DTT, and 5 µL 0.5% RapiGest were added, and the mixture was incubated at 60 °C for 30 min. Alkylation was performed by adding 2.5 µL 200 mM IAA and 5 µL 200 mM AmBic, followed by incubation in the dark for 30 min at room temperature. Proteolytic digestion was initiated with trypsin/Lys-C at an enzyme-to-protein ratio of 1:100 and the samples were incubated at 37 °C for 1 h. Subsequently, trypsin was added in 1:10 ratio and digestion continued at 37 °C for 15 h. The reaction was stopped by 0.5 µL FA and the solvents were evaporated. *N*-glycopeptides were enriched using acetone precipitation (120, 121). Dried peptide samples were dissolved in 15 µL 1% FA, followed by the addition of 150 µL ice-cold acetone. Samples were incubated at −20 °C for 18 h and centrifuged at 12000 g for 10 min at room temperature. Under these conditions, *N*-glycopeptides were predominantly retained in the pellet, while most non-glycosylated peptides remained in the supernatant. After separation, both fractions were dried down and the non-glycosylated peptide fraction was purified on C_{18} spin cartridges (106, 107).

The cartridges were equilibrated with 2 × 200 µL 50% methanol, 2 × 200 µL 0.5% TFA + 5% ACN, 2 × 200 µL 0.1% TFA solution. The samples were then applied and reapplied in 50 µL 0.1% TFA, washed with 2 × 100 µL 0.1% TFA, and eluted with 2 × 50 µL 0.1% TFA + 70% ACN. Eluates were dried and both peptide and *N*-glycopeptide fractions were stored at −20 °C.

3.2.4.2. LC-MS/MS analysis of peptides

Measurements were performed on a timsTOF HT MS (Bruker Daltonics, Germany) connected to a Dionex Ultimate 3000 nano-UHPLC system (Thermo Fisher Scientific, USA). Proteomic samples were reconstituted in 0.1% FA + 2% ACN and 200 ng of peptides were injected. Peptides were first trapped on a Thermo Fischer Acclaim PepMap C_{18} trap column (5.0 µm, 300 µm × 5 mm) and then separated on a Bruker monolithic

capillary MOSAIC C_{18} analytical column (75 µm × 150 mm) held at 50 °C. The mobile phases consisted of 0.1% FA in water (A) and 0.1% FA in 80% ACN (B). Peptides were separated using a linear gradient from 5% to 40% B over 20 min at a flow rate of 0.5 µL/min.

MS analysis was performed in positive ion mode using a CaptiveSpray 1 ion source. DIA was implemented using the DIA-parallel accumulation-serial fragmentation (DIA-PASEF) approach with a mass range of 100-1700 *m*/*z* and an ion mobility range of 0.7-1.4 V·s/cm². DIA window schemes were optimized using the py_diAID (122) tool, generating 24 non-overlapping windows covering the ion mobility-*m*/*z* space. The TIMS analyzer was operated with an accumulation time and ramp time of 180 ms. Transfer time and pre-pulse storage time were set to 60 µs and 12 µs, respectively. The total cycle time was 0.93 s and consisted of one MS1 scan followed by four dia-PASEF MS/MS scans. Precursor ions with up to 5 charges were considered for acquisition, with an intensity threshold of 1500 and a target intensity of 15000. Dynamic exclusion was enabled with a release time of 0.4 min, and exclusion windows of 0.015 *m*/*z* and 0.015 V·s/cm^2 were applied for mass and ion mobility dimensions, respectively.

3.2.4.3. Proteomic data processing and statistical analysis

Proteins were identified and quantified using DIA-NN (123) based on the Uniprot human database (accessed March 2024). Trypsin/P enzyme was used, mass accuracy was determined by the software, and match between runs was enabled. Carbamidomethylation of cysteine was a fixed modification, while methionine oxidation, *N*-terminal methionine excision, and protein *N*-terminal acetylation were variable modifications. One missed cleavage and one variable modification were allowed.

Statistics and visualization were carried out with custom code in R, assisted by RStudio. Proteins with fewer than two unique peptides, detected in at least one-third of the media samples or present in less than half of the samples in each group, were excluded. Missing values were imputed based on the number of missing values: if the respective protein was detected in less than two-thirds of the samples in the group, the 5th percentile of the sample was imputed, otherwise imputation was performed using the k-nearest neighbors algorithm (VIM package, k = 15). The data was logarithmized, and normal distribution was checked using the Shapiro-Wilk test, while equality of variances was checked using the Levene test. Based on the results, the Student's *t*-test, Welch *t*-test, or Wilcoxon rank

sum test was applied to each protein. The *p*-values were corrected using the Benjamini-Hochberg approach, and values of $p < 0.05$ were accepted as significant. Visualization consisted of PCA, clustered heatmap, and boxplots. Altered biological processes were explored using gene set enrichment analysis (GSEA, clusterProfiler package, *p*-value threshold: 0.1).

3.2.4.4. LC-MS/MS analysis of *N*-glycopeptides

Samples were measured on an Exploris 240 Orbitrap (Thermo Fisher Scientific, USA) coupled to a Waters nanoAcquity (Milford, USA). *N*-glycopeptides were dissolved in 10 µL 0.1% FA + 2% ACN and 2 µL was injected. Chromatographic separation was achieved using a Waters Symmetry C_{18} trap column (5 µm, 180 µm × 20 mm) and a Waters Acquity M-Class BEH130 C_{18} analytical column (1.7 µm, 75 µm × 250 mm) heated to 50 °C. Mobile phases consisted of 0.1% FA in water (A) and 0.1% FA in ACN (B), delivered at a flow rate of 300 nL/min. The gradient started at 2% B for 1 min, then increased to 25% B over 80 min, 40% B over 3 min, and 90% B over 1 min, held for 2 min, then equilibrated to 2% B for 2 min.

A nanospray ion source was used in positive mode, with a capillary voltage of 1.8 kV and a capillary temperature of 275 °C. Full MS scans were collected within the 360-2200 *m*/*z* range at a resolution of 120000, with an automatic gain control target value of $2 \cdot 10^6$ and a maximum injection time of 200 ms. Selected precursor ions were isolated within a 2 Da window and stepwise higher energy collisional dissociation (10, 20, and 30 eV) was used to collect MS/MS in the 200-2000 *m*/*z* range.

3.2.4.5. *N*-glycoproteomic data processing and statistical analysis

Proteins included in the statistical evaluation of the proteomics dataset were filtered based on the presence of known *N*-glycosylation sites. Only proteins with confirmed *N*-glycosylation were retained, resulting in a set of 214 proteins. The *N*-glycan compositions obtained from *N*-glycomic measurements were normalized to total ion intensity and averaged between A549 and BEAS-2B sEV samples. Only *N*-glycans with an average relative abundance of at least 0.25% were considered for further analysis, and biosynthetically unlikely compositions were removed. After converting the derivatized *N*-glycans to their non-derivatized forms, a final list of 34 *N*-glycan compositions was compiled. Identification and quantification of *N*-glycopeptides were performed using GlycReSoft (124), applying the protein and *N*-glycan databases described above. The

MS1 mass tolerance was 10 ppm, while the MS/MS mass tolerance was 20 ppm, and maximum 2 missed cleavages were allowed. Carbamidomethylation of cysteine was a fixed modification, while oxidation of methionine was a variable modification.

Further data processing was carried out in R. *N*-glycopeptide identifications were accepted if their MS1 score exceeded 3 and the MS2 score exceeded 5 (125). Signals originating from cell culture media components were removed in the same manner as described for the proteomics dataset, and *N*-glycopeptides assigned to unknown *N*-glycosylation sites were excluded. Signal intensities were normalized using total area normalization. Statistics and visualization of the *N*-glycoproteomics dataset followed the same workflow as applied for the proteomic data. To further characterize *N*-glycosylation patterns, derived metrics were calculated, including the proportion of sialylated antennae, the average rate of fucosylation per *N*-glycopeptide, and the relative abundance of different *N*-glycopeptide classes.

3.2.5. CS/DS glycosaminoglycan analysis

For CS/DS digestion, 10 µg protein samples were diluted to 70 µL with water, then 20 µL 500 mM AmBic solution, 5 µL 100 mM ammonium acetate, and 5 µL 5 mU/µL chondroitinase-ABC were added, and incubated at 37 °C for 16 h. Digestion was stopped by heating at 90 °C for 3 min, then the samples were dried down and kept at −20 °C. The purification of CS/DS disaccharides was performed in the same manner as described for tissues.

The measurement of CS/DS disaccharides was performed as described above for tissue samples, with the following modifications. Samples were dissolved in 8 µL 10 mM ammonium formate + 75% ACN (pH = 4.4) and 2 µL was injected. Chromatographic separation was carried out at a flow rate of 10 µL/min using an isocratic elution at 5% B for 7 min, followed by column washing with 95% B for 5 min and equilibration at 5% B for 3 min. MS measurements were performed with a capillary voltage of 2.5 kV and a cone voltage of 10 eV. The transfer collision energy was set to 4 eV for MS1 and 20 eV for MS/MS.

The integration of CS/DS disaccharide peaks and the calculation of their relative abundances were performed using the same procedure described above for tissue samples. In one sample, disaccharides were not detectable, probably due to a sample preparation issue; therefore, this sample was removed from further analysis. Data processing,

visualization, and statistical analyses were carried out in R with custom scripts. Statistical significance was assessed using Student's *t*-test, Welch's *t*-test, or the Wilcoxon rank-sum test, and Benjamini-Hochberg correction was performed. Data visualization included boxplots, PCA, and clustered heatmap.

4. Results

4.1. Tissue study

In the tissue study, proteomic and GAG (CS/DS, HS) disaccharide profiles of 22 regions from seven ALK rearranged lung cancer patients were analyzed. The tumor areas were categorized by pathologists according to morphology (solid tumor, $n = 10$; tubular tumor, $n = 3$; papillary tumor, $n = 5$), as well as mucin (moderate, M2, $n = 4$; high, M3, $n = 5$) and stromal (low, S1, $n = 5$; moderate, S2, $n = 5$; high, S3, $n = 7$) content. In addition to tumor regions, adjacent normal lung tissue regions (ANL, $n = 4$) were also included in the analysis.

4.1.1. Proteomic analysis

3411 proteins were identified from proteomic measurements (an average of 651 proteins per region), of which 2092 proteins were quantified. Differential expression analysis revealed substantial quantitative differences between tissue regions with different morphological classifications. The number of dysregulated proteins observed in each pairwise comparison and the distribution of upregulated and downregulated proteins are presented in **Table 2**.

Table 2. Number of differentially expressed proteins identified in pairwise comparisons of tissue regions. Upregulated and downregulated proteins are shown separately.

	Dysregulated proteins	Upregulation	Downregulation
solid - ANL	541	369	172
papillary - ANL	380	250	130
tubular - ANL	250	158	92
papillary - solid	199	66	133
tubular - solid	83	22	61
papillary - tubular	40	27	13

Certain proteins changed in the same direction in all tumor sample groups, for example, the ERGIC-53 protein was overexpressed (fold change (FC) = 1.5-2.2), while stomatin was underexpressed in all tumor regions (FC = 0.26-0.38) compared to adjacent normal regions. Other proteins, such as lactotransferrin and mucin-5B, showed different expression patterns across different tumor types. Representative boxplots illustrating these expression differences are shown in **Fig. 5**.

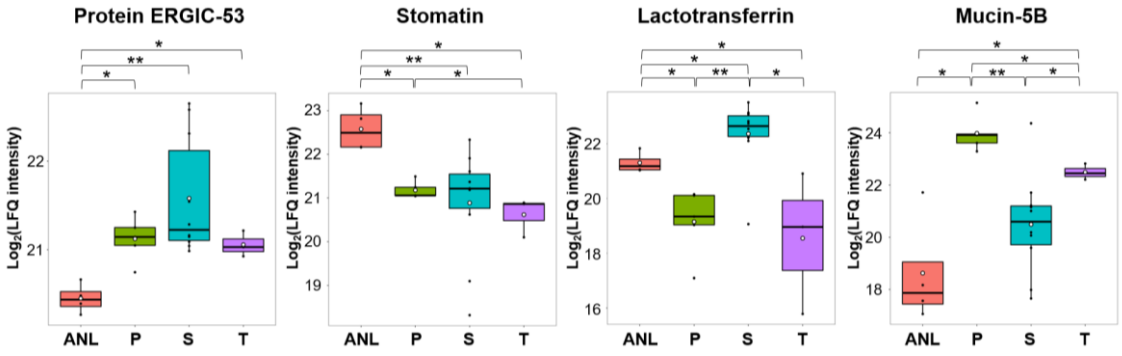

Figure 5. Selected differentially expressed proteins between lung tissue regions with different morphological classifications. (*$p < 0.05$ (or 0.057 for ANL vs T); **$p < 0.01$).

Since GAG chains are covalently bound to PG core proteins, we specifically examined the proteins belonging to this group. Among the characterized proteins, three HSPG core proteins were found to be differentially expressed between adjacent normal and tumor tissue regions, while no significant differences were observed between different tumor subtypes. Agrin and perlecan were consistently downregulated in all tumor regions compared to adjacent normal lung tissue (FC = 0.27-0.55, see **Fig. 6**). In addition, the collagen alpha-1(VIII) chain showed reduced expression in solid tumor regions (FC = 0.37).

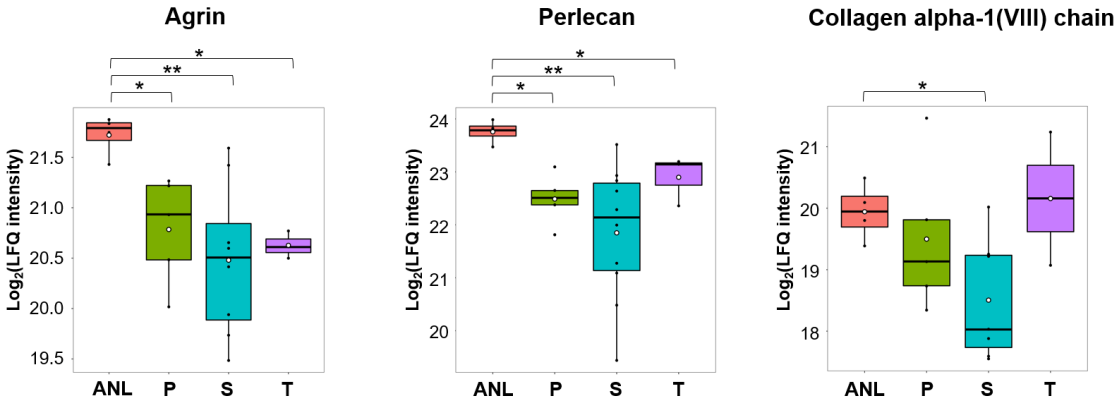

Figure 6. Differentially expressed proteoglycan core proteins between lung tissue regions with different morphological classifications (*$p < 0.05$ (or 0.057 for ANL vs T); **$p < 0.01$).

To explore the biological functions associated with the differentially expressed proteins, functional enrichment analyses were performed. In **Fig. 7**, each protein is represented by a node, while the connections between proteins are indicated by edges. When comparing different tumor morphologies, the dysregulated proteins were predominantly associated

with protein synthesis-related processes. For example, in the comparison between papillary and solid tumor regions, two major clusters were identified corresponding to RNA splicing and translation (**Fig. 7/A**). Proteins belonging to these clusters were generally less abundant in papillary tumors than in solid tumors.

In contrast, comparisons between adjacent normal lung tissue and tumor regions revealed a broader range of altered biological processes and cellular localizations. For instance, the network generated from proteins dysregulated between adjacent normal and solid tumor regions contained clusters associated with mitochondrial proteins, protein folding, ECM organization, translation, and RNA splicing (**Fig. 7/B**). In this comparison, proteins involved in ECM organization were mainly decreased in tumor regions, while proteins associated with mitochondrial functions, protein folding, and protein synthesis-related processes were generally increased relative to adjacent normal tissue.

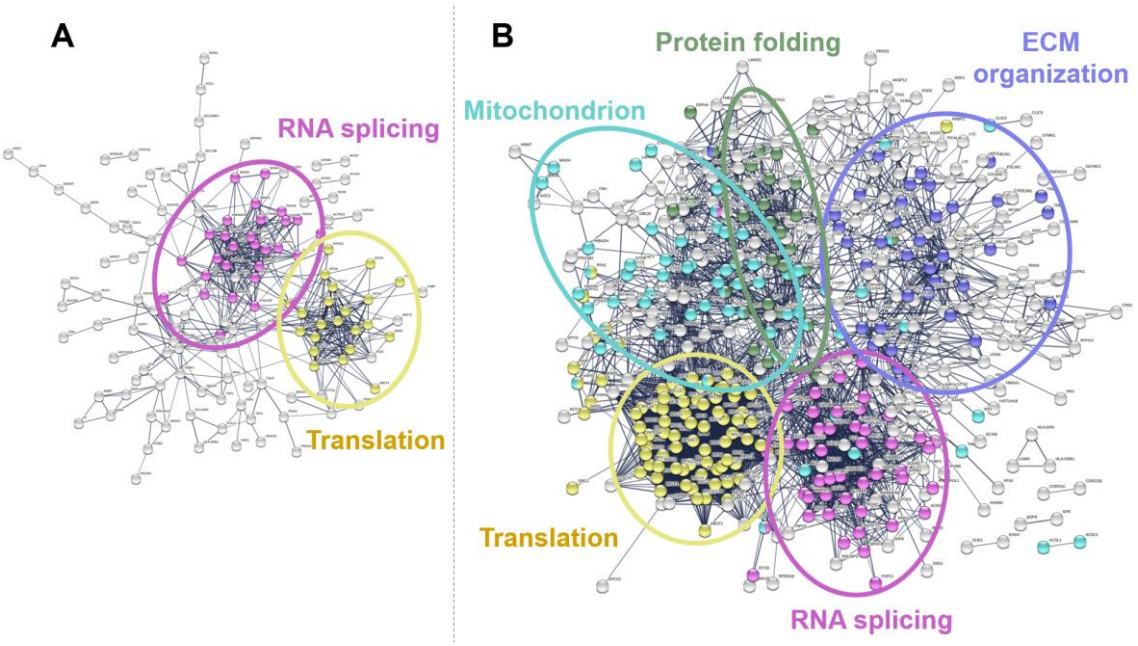

Figure 7. Protein interaction networks on proteins dysregulated between papillary and solid tumor (**A**) and solid tumor and adjacent normal lung regions (**B**).

To visualize global proteomic differences between the investigated tissue regions, clustered heatmap was performed for proteins quantified in more than 9 regions (**Fig. 8/A**) and PCA analysis was performed for proteins quantified in all regions (**Fig. 8/B-C**). PC1 and PC2 explained 34.5% and 14.4% of the variance in the data, respectively. Both methods showed a clear separation between ANL samples and tumor regions, with ANL samples clustering closely together. Among the morphological

subtypes, most solid tumor regions (8 out of 10) grouped within the same cluster, indicating relatively similar proteomic characteristics. In contrast, papillary tumor regions formed two subgroups, which corresponded to differences in mucin content. The tubular tumor regions showed greater variability and they did not cluster together.

In terms of mucin content, regions classified as M2 or M3 (which were primarily associated with tubular and papillary tumors) formed partially separated groups. In contrast, stromal content did not result in clear grouping, and regions with similar stromal content were in different clusters. Further results regarding proteomic differences associated with mucin and stromal content are discussed in (104).

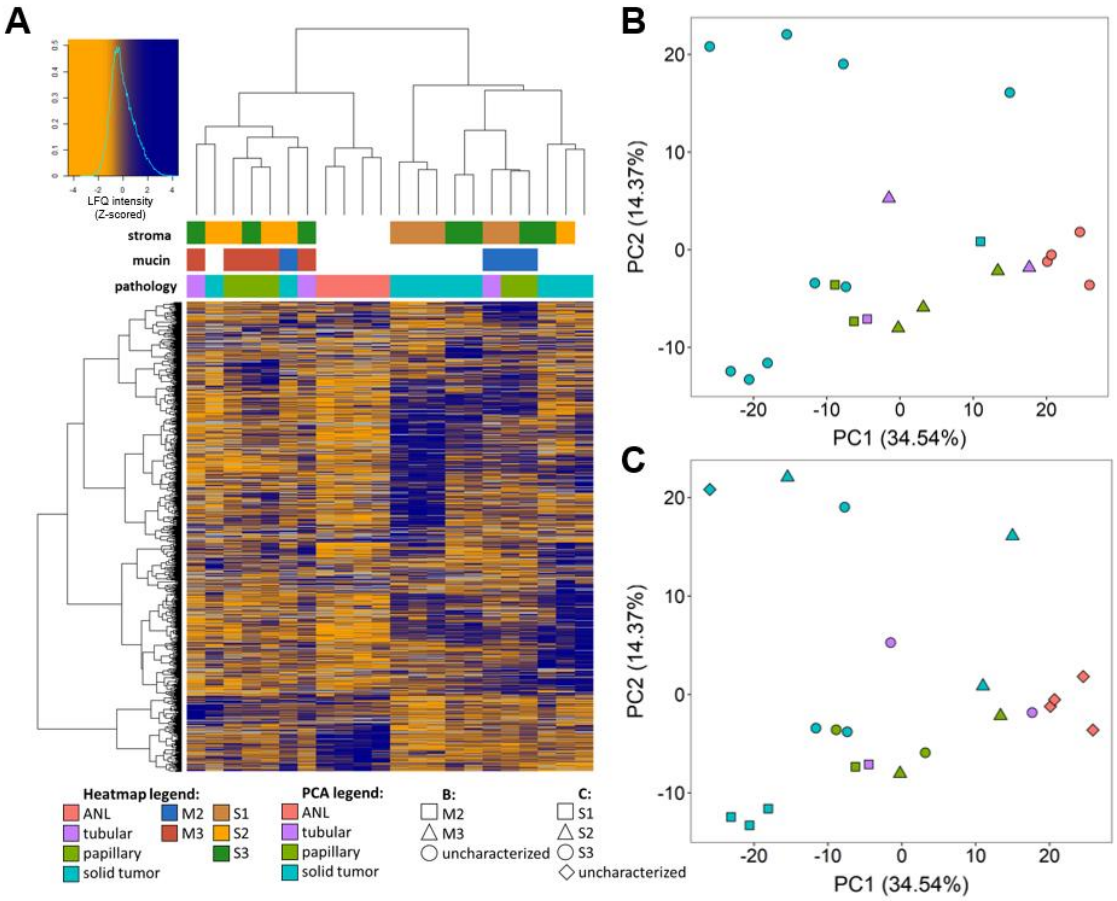

Figure 8. Proteomic clustered heatmap (**A**) and principal component analyses (**B, C**). (**A**) was performed for proteins quantified in more than 9 regions, while (**B, C**) was performed for proteins quantified in all regions. In PCA, samples are labelled based on morphological classification and mucin (**B**) or stromal content (**C**).

4.1.2. Glycosaminoglycan analysis

In the GAG analysis, all four CS/DS disaccharides and all six HS disaccharides could be identified and quantified in all samples. The ANL regions contained the lowest average amount of CS/DS disaccharides, while the total CS/DS amount increased by 4.4-, 4.0-, and 2.5-fold on average in the tubular, papillary, and solid tumor regions, respectively (**Fig. 9/A**). In all sample groups, the most common CS/DS disaccharide was D0a0, but in tumor samples, the proportion of sulfated disaccharides increased by FC = 1.4-7.5 (**Fig. 9/B**).

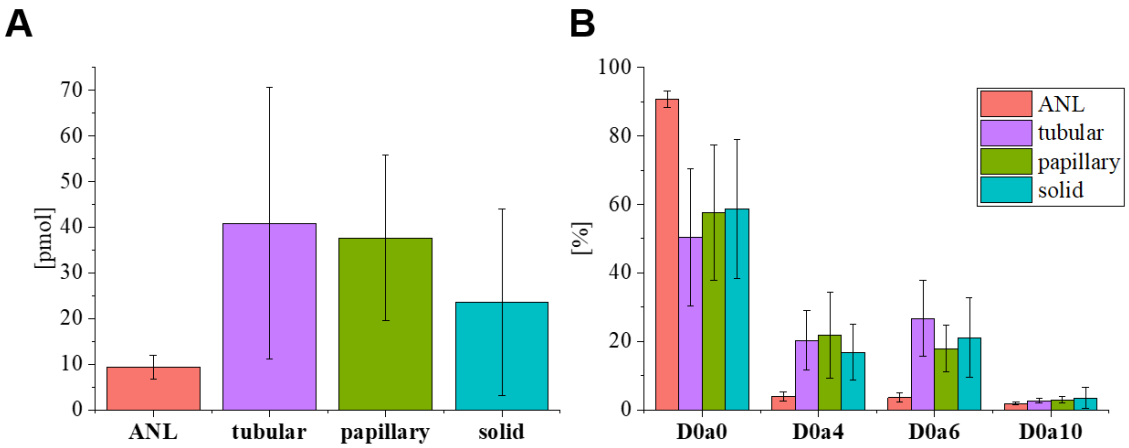

Figure 9. Total amount (**A**) and relative abundance (**B**) of CS/DS disaccharides in lung tissue regions with different morphological classifications.

Next, we compared two derived values: the average rate of CS/DS sulfation and the ratio of 6S/4S sulfated components. In ANL regions, the CS/DS disaccharides carried an average of 0.11 sulfate groups, while in tumor regions this value increased to 0.45-0.52 (**Fig. 10/A**). In the ANL and papillary tumor regions, 4S sulfation was slightly overrepresented (6S/4S ratio slightly below 1), while in the tubular and solid tumor regions, 6S sulfation dominated (6S/4S ratio between 1.2 and 1.3, **Fig. 10/B**).

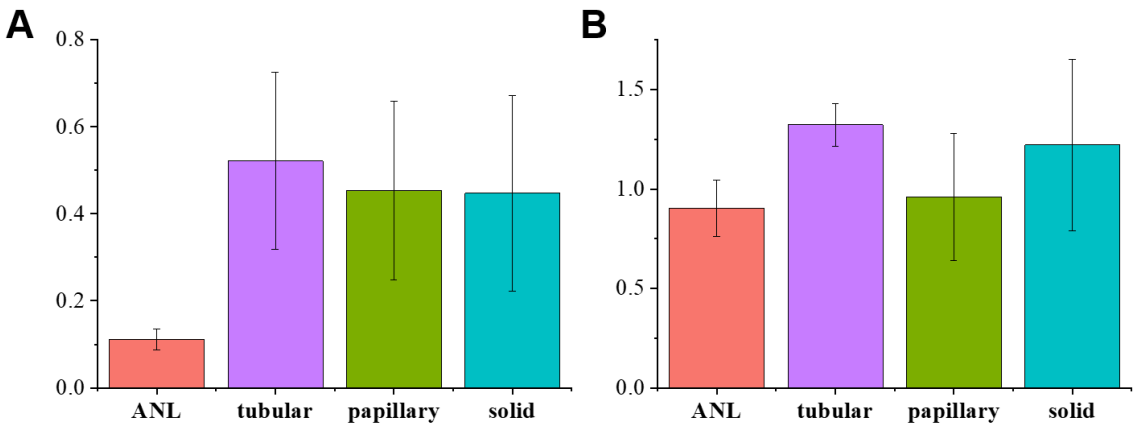

Figure 10. Average rate of CS/DS sulfation (**A**) and ratio of 6S/4S sulfated CS/DS disaccharides (**B**) in lung tissue regions with different morphological classifications.

The lowest amounts of HS disaccharides were found in the ANL regions, while the tubular, papillary, and solid morphological regions contained 8.5, 4.8, and 2.5 times more HS, respectively. The total amount of HS disaccharides in each group showed the same trend as the total amount of CS/DS disaccharides, but in the case of HS, greater differences were observed (**Fig. 9/A** vs **Fig. 11/A**). In ANL regions, D0A0 disaccharide was the most common (84%), while in tumor regions, its proportion decreased to 40-54% (**Fig. 11/B**). In contrast, the amount of D2A0 + D0A6 disaccharides increased from 7% to 42-56% in tumor regions. Among the less abundant disaccharides, the frequency of D2A6 increased, while the amount of D0S0, D2S0 + D0S6, and D2S6 disaccharides decreased in the tumor sample groups compared to ANL.

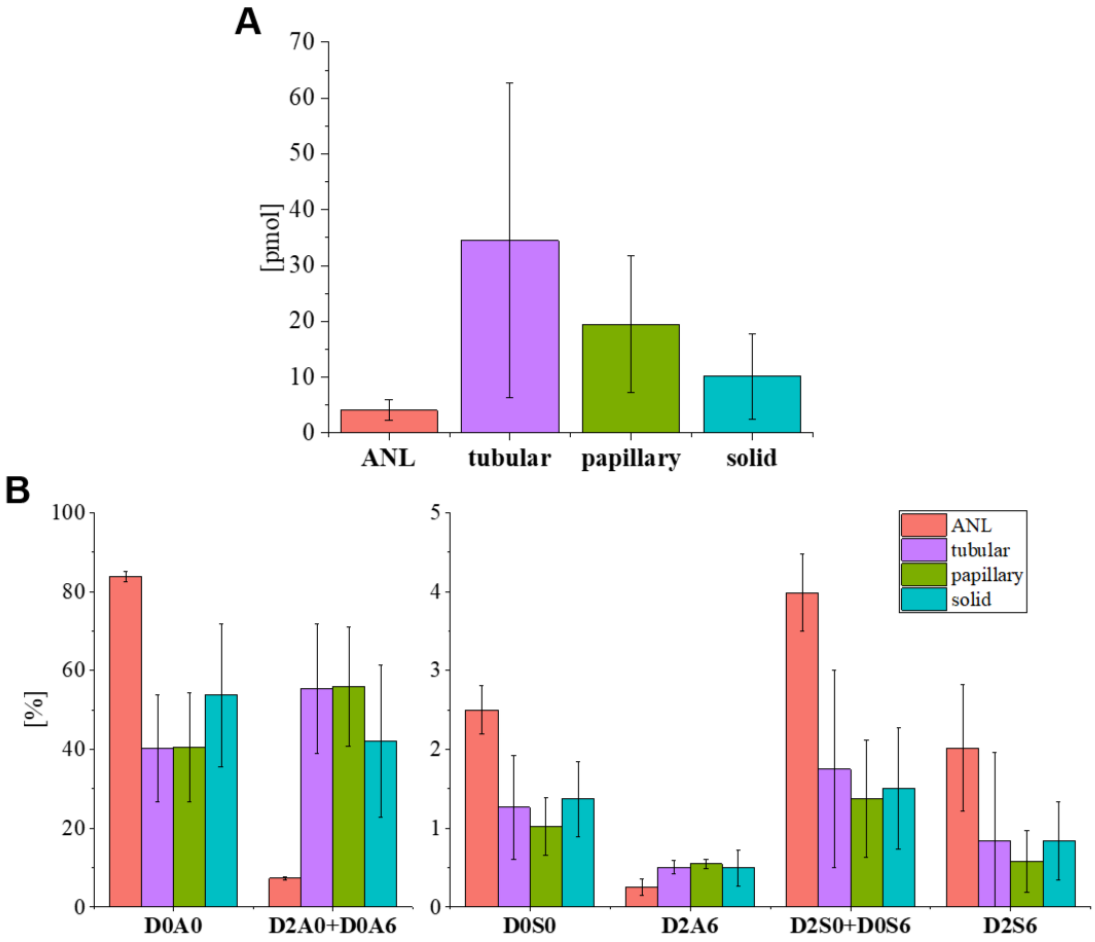

Figure 11. Total amount (**A**) and relative abundance (**B**) of HS disaccharides in lung tissue regions with different morphological classifications.

HS disaccharides in the ANL regions had 0.24 sulfate groups on average, while this rose to 0.50-0.64 in different tumor types (**Fig. 12/A**). The rate of *N*-sulfation decreased markedly in tumor areas compared to adjacent normal tissue (**Fig. 12/B**). This change was primarily the result of differences observed in monosulfated disaccharides (FC = 0.06-0.19, **Fig. 12/C**) and, to a lesser extent, in disulfated disaccharides (FC = 0.59-0.66, **Fig. 12/D**).

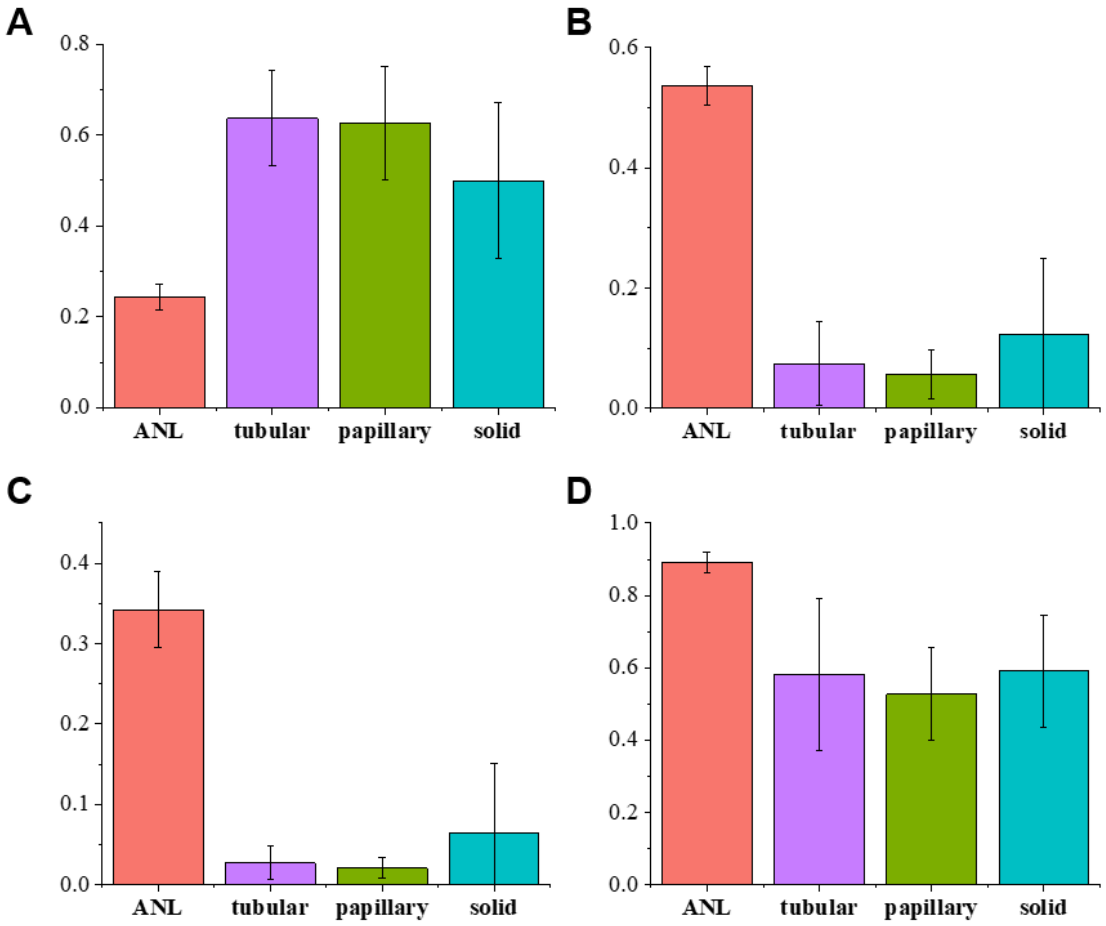

Figure 12. Average rate of HS sulfation (**A**) and ratio of *N*/*O*-sulfation for all (**B**), monosulfated (**C**) and disulfated HS disaccharides (**D**) in lung tissue regions with different morphological classifications.

Clustered heatmap (**Fig. 13/A**) and PCA analyses (**Fig. 13/B-C**) were used to visualize overall changes observed on the GAG levels. PC1 accounted for 66.2%, while PC2 accounted for 16.6% of the variance. In both analyses, the ANL regions were well separated from the tumor samples; however, one solid tumor sample showed a GAG profile similar to that of ANL. Papillary tumor regions split into two subgroups based on their mucin content, just like in case of proteomics. Tubular and solid tumor regions showed high levels of variation and did not cluster together.

In terms of mucin content, the M2 and M3 regions formed completely distinct groups (**Fig. 13/A**). In contrast, regions with different stromal content did not cluster together. For further findings regarding differences in GAG composition associated with mucin and stromal content, see (104).

In the clustering of disaccharides (**Fig. 13/A**), one of the main groups consists of all *O*-sulfated disaccharides (D0a4, D0a6, and D0a10 CS/DS disaccharides, as well as the D2A0 + D0A6 and D2A6 HS disaccharides), while the other consisted of non-sulfated disaccharides (D0a0, D0A0) and *N*-sulfated HS disaccharides (D0S0, D2S0 + D0S6, D2S6). The first cluster was typically overrepresented in tumor regions compared to ANL, while the second one was generally underrepresented.

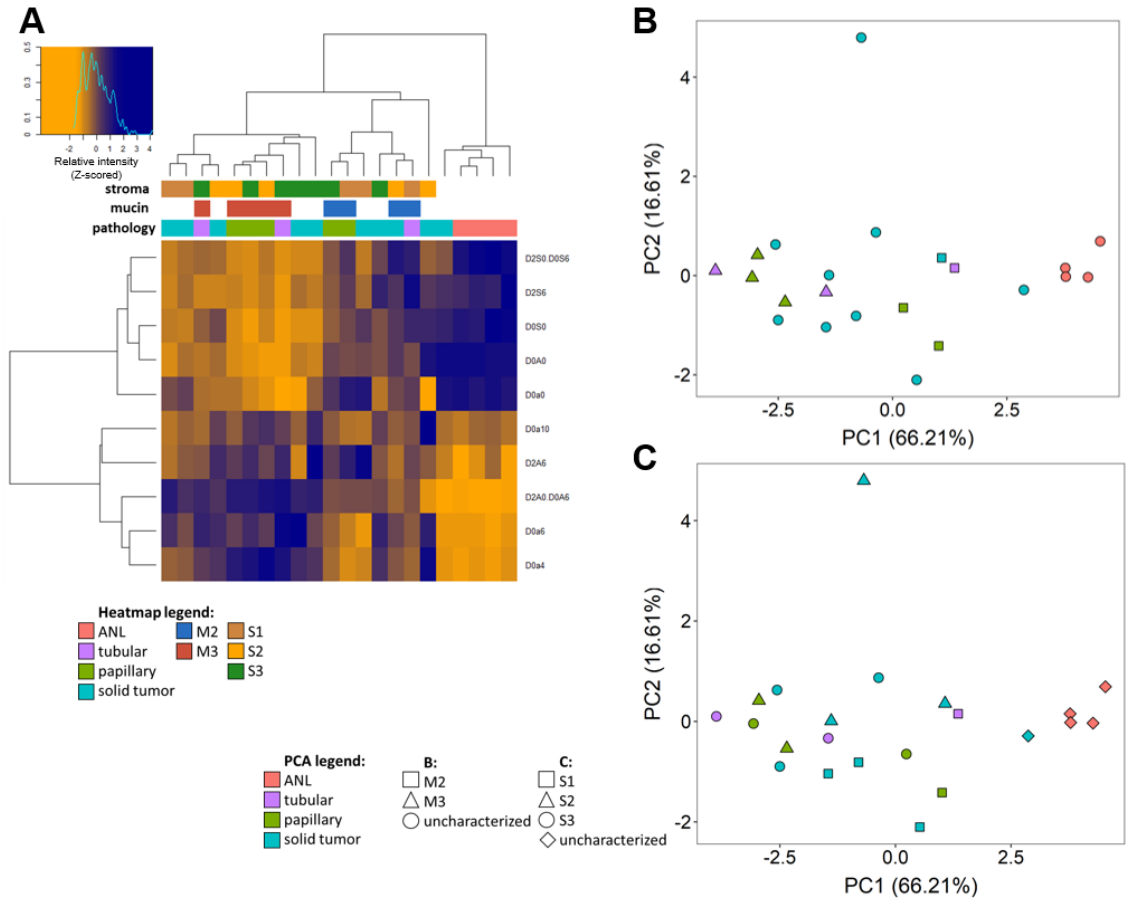

Figure 13. GAG clustered heatmap (**A**) and principal component analyses (**B, C**). In PCA, samples are labelled based on morphological classification and mucin (**B**) or stromal content (**C**).

4.2. Extracellular vesicle study

In the EV study, proteomic, *N*-glycoproteomic, and CS/DS GAG profiles of sEVs and the corresponding medium controls were analyzed. Six replicates were prepared from both A549 and BEAS-2B sEVs, while three replicates were used from each control medium to filter out medium-derived contaminants.

4.2.1. Characterization of small extracellular vesicles

Isolates were characterized according to the Minimal information for studies of extracellular vesicles 2023 guidelines (97). Particle shapes were accessed using TEM, while size distribution was measured by MRPS. **Fig. 14/A** and **B** show results for A549 and BEAS-2B sEVs, respectively. TEM analysis showed that spherical particles were present in the samples and were generally smaller than 200 nm. The size of the A549 particles, as determined by TEM, was typically 30-120 nm, while the BEAS-2B particles were larger and exhibited a broader size distribution (50-250 nm, see **Fig. 14/C**).

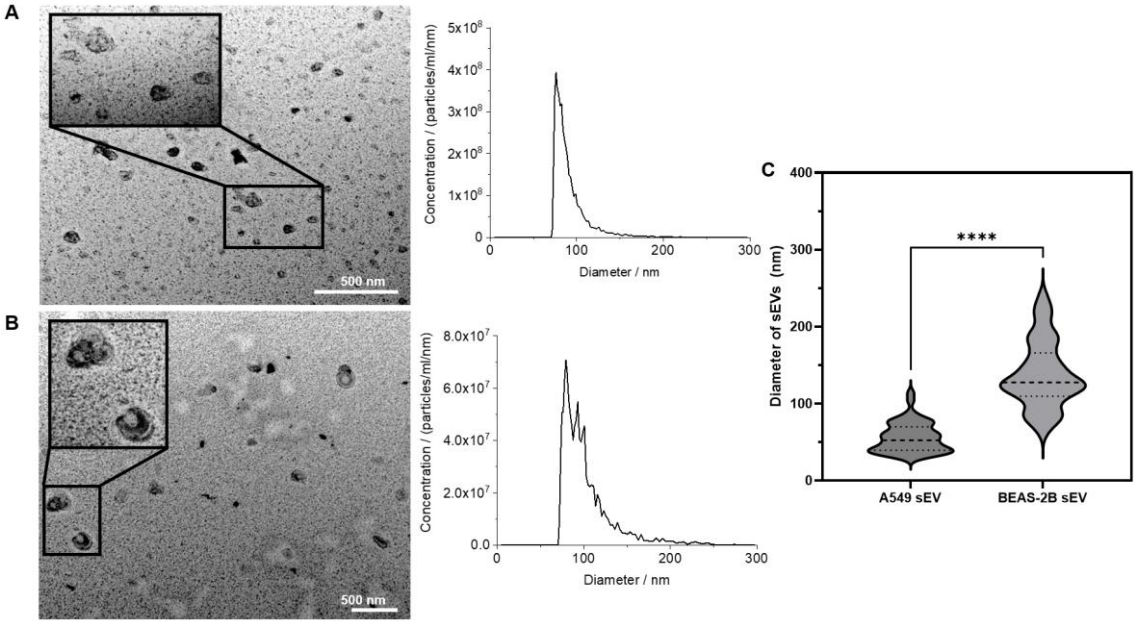

Figure 14. Characterization of A549 (**A**) and BEAS-2B (**B**) sEVs with transmission electron microscopy and microfluidic resistive pulse sensing. (**C**) illustrates sEV diameters measured by TEM (N = 50-50 particles). (****$p < 0.0001$)

Compared to TEM, MRPS analysis showed smaller particle sizes. At the same time, MRPS confirmed that BEAS-2B sEVs were larger, with approximately 40% of the particles being larger than 100 nm and some particles even exceeding 150 nm, whereas

in the case of A549, only 20% of the particles were larger than 100 nm and particles larger than 150 nm were not common.

In addition, MRPS analysis showed that sEVs were present in higher concentrations in the A549 sample than in the BEAS-2B sample. Therefore, in the following steps, the same amounts of proteins were used to minimize the differences.

4.2.2. Proteomic analysis

2334 proteins were quantified from proteomic measurements of the 12 sEV and 6 media samples. One of the most used reference lists for characterizing the quality of EV proteins is the Exocarta Top 100 EV protein list. Compared to this, all sEV samples contained 77-91 of the 100 most frequently identified EV proteins, confirming that the isolates were consistent with those described in the literature.

After filtering out proteins identified by a single peptide, identified in limited number of samples, or found in culture media, 945 proteins were subjected to statistical analysis. A significant portion of the hits, about one-third of the proteins, were excluded because they were also present in the culture media, primarily in BEGM. Out of the 945 proteins quantified, 408 proteins showed different expressions between A549 and BEAS-2B sEVs. 313 proteins were upregulated, most notably alpha-fetoprotein and aldehyde dehydrogenase 3 family member A. On the other hand, 95 proteins were downregulated, including pentraxin-related protein PTX3 and fibulin-2.

Eleven PG core proteins were included in statistical analysis, of which seven showed differential expression: versican core protein (CS), chondroitin sulfate proteoglycan 4 (CS), aggrecan core protein (CS/KS), testican-1 (HS/CS), syndecan-4 (HS/CS), collagen alpha-1(XVIII) chain (HS), and mimecan (KS). The differences in CSPG expression are presented in **Fig. 15**. These proteins showed different trends: versican and testican-1 were upregulated, while CSPG4, aggrecan, and syndecan-4 were downregulated in A549 sEVs compared to BEAS-2B sEVs.

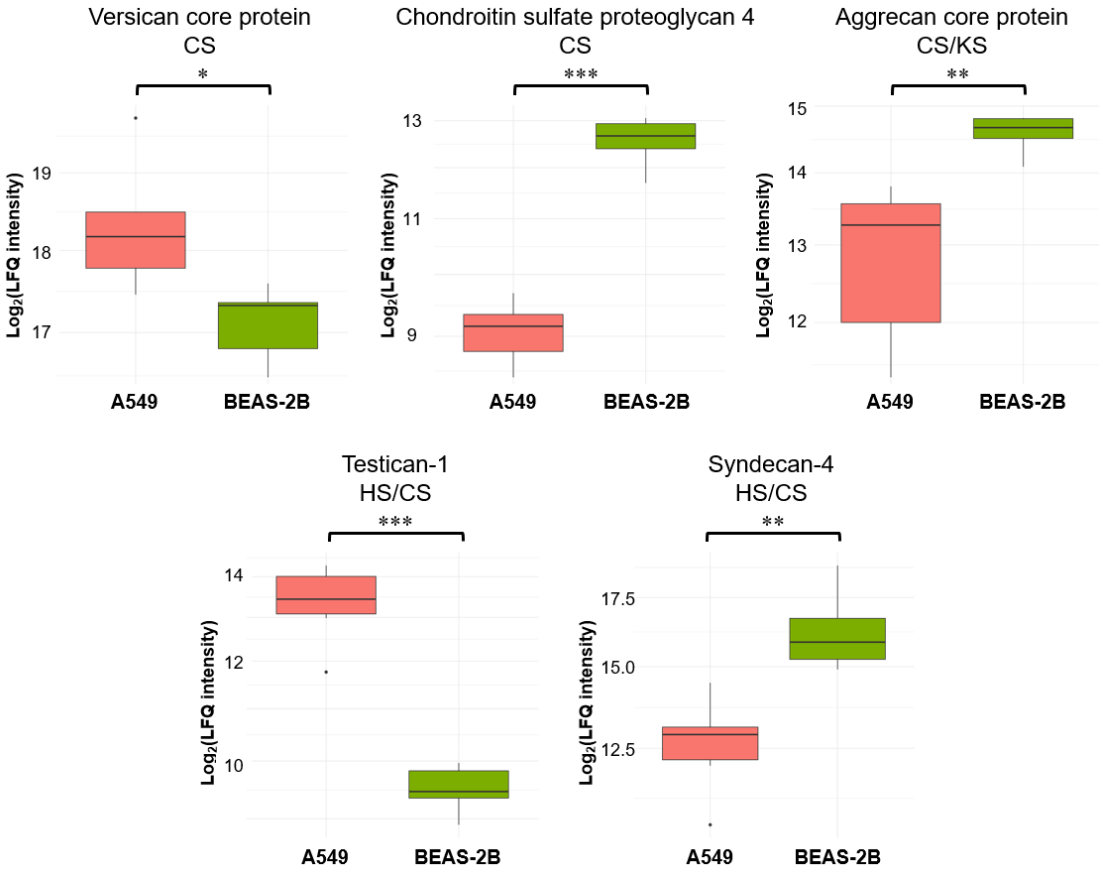

Figure 15. Differentially expressed CS proteoglycan core proteins between A549 and BEAS-2B sEVs. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$)

GSEA was performed to identify altered biological processes between A549 and BEAS-2B sEVs. A total of 60 enriched pathways were identified and characterized by normalized enrichment scores (NES), indicating the up- or downregulation of the given process. Most strongly enriched biological process was negative regulation of cell cycle processes (NES = 1.75). Other upregulated terms included cell cycle checkpoint signaling, cellular nitrogen compound metabolism, and regulation of DNA metabolic processes. In addition, several processes associated with nucleic acid metabolism, protein synthesis, and broader biosynthetic and metabolic pathways showed significant enrichment. Conversely, a few biological processes displayed negative enrichment, suggesting reduced representation in A549 sEVs. These included processes related to immune response, immune effector process, and positive regulation of endocytosis.

Finally, PCA and clustered heatmap analyses were used to characterize differences between the proteomic profiles. **Fig. 16/A** shows the PCA for all proteins, while **Fig. 16/B** shows the clustered heatmap for differentially expressed proteins. Both methods

separated A549 and BEAS-2B sEVs completely. In PCA analysis, the separation was based on PC1, which explained 42.3% of the variance. This differentiation also confirmed the significantly altered proteomic profiles between the two sEV groups, as indicated by the high number of differentially expressed proteins.

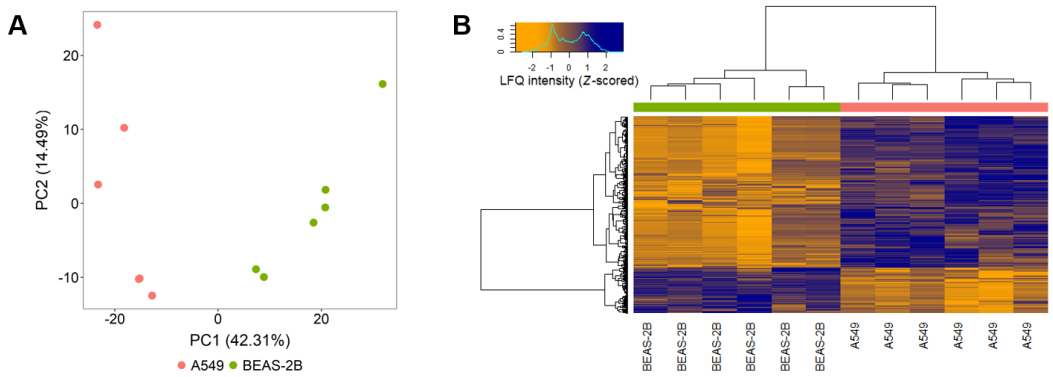

Figure 16. Proteomic principal component analysis (**A**) and clustered heatmap (**B**). (**A**) was performed for all quantified proteins, while (**B**) was performed for differentially expressed proteins only.

4.2.3. Glycomics guided *N*-glycoproteomic analysis

Prior to *N*-glycoproteomic analysis, released *N*-glycan profiles were first characterized to establish an *N*-glycan library for subsequent *N*-glycopeptide identification. In total, 271 *N*-glycan structures were detected in at least one of the samples analyzed. The *N*-glycan dataset also included phosphorylated structures and highly fucosylated species; however, these *N*-glycans were generally present at low relative abundance, suggesting that they are less likely to be detected in *N*-glycoproteomics. Among the sEV samples, 129 *N*-glycan compositions were identified in A549 sEVs and 203 in BEAS-2B sEVs. In contrast, only 5 and 13 *N*-glycans were detected in the F12 and BEGM culture media, respectively. These results confirmed that the *N*-glycan profiles observed in the sEV samples predominantly originated from the sEVs themselves rather than from culture medium contamination.

In *N*-glycoproteomics, 1659 *N*-glycopeptides were identified and quantified, which were assigned to 162 proteins and 793 peptides. However, a significant proportion of the *N*-glycopeptides could not be reliably identified in most samples, and the ratio of missing values was nearly 80%. Thus, after filtering, 227 *N*-glycoforms were left for statistical analysis. These *N*-glycoforms belonged to 117 *N*-glycosylation sites of 72 proteins.

Among these, 152 *N*-glycoforms were differentially represented between the two sEV types: 92 were overrepresented, while 60 were underrepresented in A549 sEVs.

12 *N*-glycoforms of versican were included in statistics, of which 8 were differentially represented (see **Table 3**). *N*-glycosylation was typically upregulated at positions 330, 1442, and 3369, while it was reduced at position 1898. Other highly *N*-glycosylated proteins included the galectin-3-binding protein, which showed different representation in 21 of its 30 *N*-glycoforms, and laminins, which showed different representation in 23 of their 29 *N*-glycoforms.

Table 3. List of versican *N*-glycoforms included in statistical analysis. A $p < 0.05$ was considered significant.

***N*-glycosylation site**	***N*-glycan**	***p***	**$\log_2$FC**
330	Fuc:1; Hex:5; HexNAc:4; Neu5Ac:1	0.001	7.13
330	Fuc:1; Hex:5; HexNAc:4; Neu5Ac:2	0.058	3.95
330	Fuc:1; Hex:5; HexNAc:4	0.008	4.63
1442	Fuc:1; Hex:5; HexNAc:4; Neu5Ac:1	0.012	4.26
1442	Fuc:1; Hex:5; HexNAc:4; Neu5Ac:2	0.057	2.58
1898	Fuc:1; Hex:7; HexNAc:6; Neu5Ac:1	0.310	-0.79
1898	Fuc:1; Hex:7; HexNAc:6	0.018	-2.32
1898	Fuc:1; Hex:8; HexNAc:7; Neu5Ac:1	0.033	-3.62
1898	Fuc:1; Hex:8; HexNAc:7	0.033	-3.86
1898	Fuc:2; Hex:6; HexNAc:5; Neu5Ac:1	0.860	-2.12
3369	Fuc:1; Hex:5; HexNAc:4	0.021	1.89
3369	Fuc:1; Hex:6; HexNAc:5	0.000	5.13

Alterations observed at the *N*-glycopeptide level may arise from changes in protein abundance, modifications in *N*-glycosylation patterns, or a combination of both effects. As an example, the versican core protein (CSPG2) *N*-glycoform carrying F1H5N4 *N*-glycan (one Fuc, five Hex, and four HexNAc) at site N3369 had a FC of 3.7, whereas the F1H6N5 *N*-glycoform detected at the same site showed a considerably higher FC of 35.0. Since the abundance of the versican itself also increased (FC = 2.3), the moderate increase observed in the F1H5N4 *N*-glycoform was largely attributable to the elevated protein level, while the more substantial increase in the F1H6N5 form suggested changes at the *N*-glycan level.

To separate the contribution of *N*-glycosylation changes from protein-level effects, *N*-glycopeptide FCs were normalized using the corresponding protein FCs obtained from

the proteomic dataset. The normalized FC values are illustrated in **Fig. 17**, where small bubbles indicate primarily changes in protein abundance, while larger bubbles represent changes in *N*-glycosylation patterns; the red and blue colors indicate over- and underrepresentation, respectively.

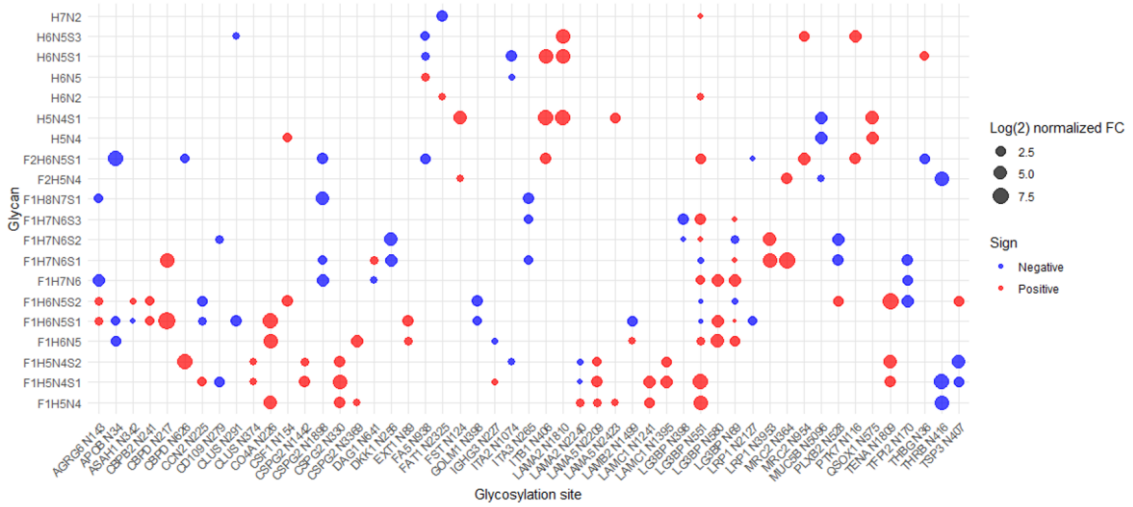

Figure 17. Bubble plot of protein intensity-normalized FCs of *N*-glycopeptides. *N*-glycosylation sites linked to at least two *N*-glycans and *N*-glycans associated with at least five *N*-glycosylation sites are presented. F, H, N, and S denote Fuc, Hex, HexNAc, and Neu5Ac, respectively.

As depicted in **Fig. 17**, laminin proteins (LAMA2, LAMA5, and LAMC1) typically carried specific *N*-glycans in increased proportions in A549 sEVs. In contrast, prothrombin (THRB) and tissue factor pathway inhibitor 2 (TFPI2) showed a general reduction in *N*-glycosylation occupancy. The integrins revealed mixed changes: the proportion of *N*-glycans generally decreased for ITA2 and ITA3, while it increased for ITB1.

Similar to proteomics, PCA (**Fig. 18/A**, all *N*-glycoforms) and clustered heatmap analyses (**Fig. 18/B**, differentially represented *N*-glycoforms) were used to visualize differences at the global *N*-glycoproteome level. Both approaches showed complete separation of A549 and BEAS-2B sEVs and confirmed the significantly altered *N*-glycosylation profiles. In the PCA, separation occurred along PC1, which explained 54.2% of the variance.

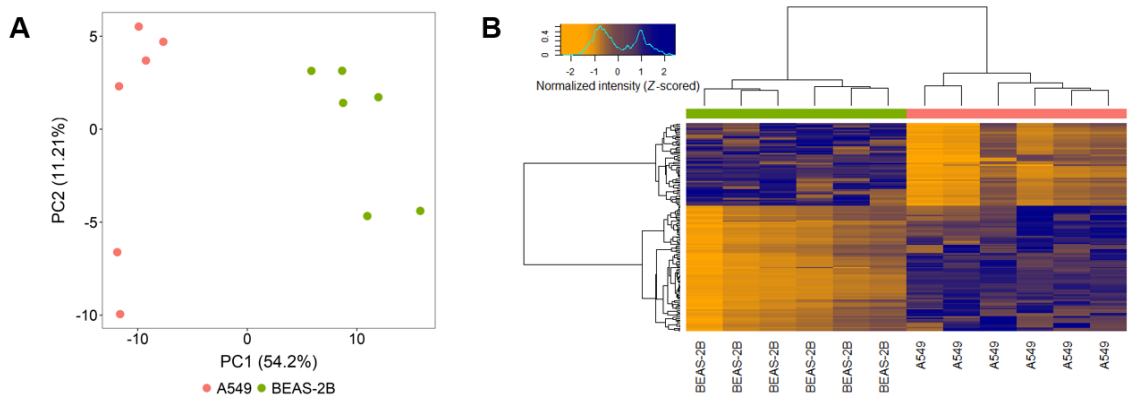

Figure 18. *N*-glycoproteomic principal component analysis (**A**) and clustered heatmap (**B**). (**A**) was performed for all quantified *N*-glycopeptides, while (**B**) was performed for differentially represented *N*-glycopeptides only.

To provide a more comprehensive picture of the observed *N*-glycosylation changes, the *N*-glycopeptide-level data was summarized into general *N*-glycosylation characteristics, analogous to those that can be obtained from the analysis of released *N*-glycans. In this approach, *N*-glycans were categorized based on their structural class (complex, hybrid, or oligomannose) and according to the extent of fucosylation and sialylation. Changes in these characteristics are shown on boxplots in **Fig. 19**.

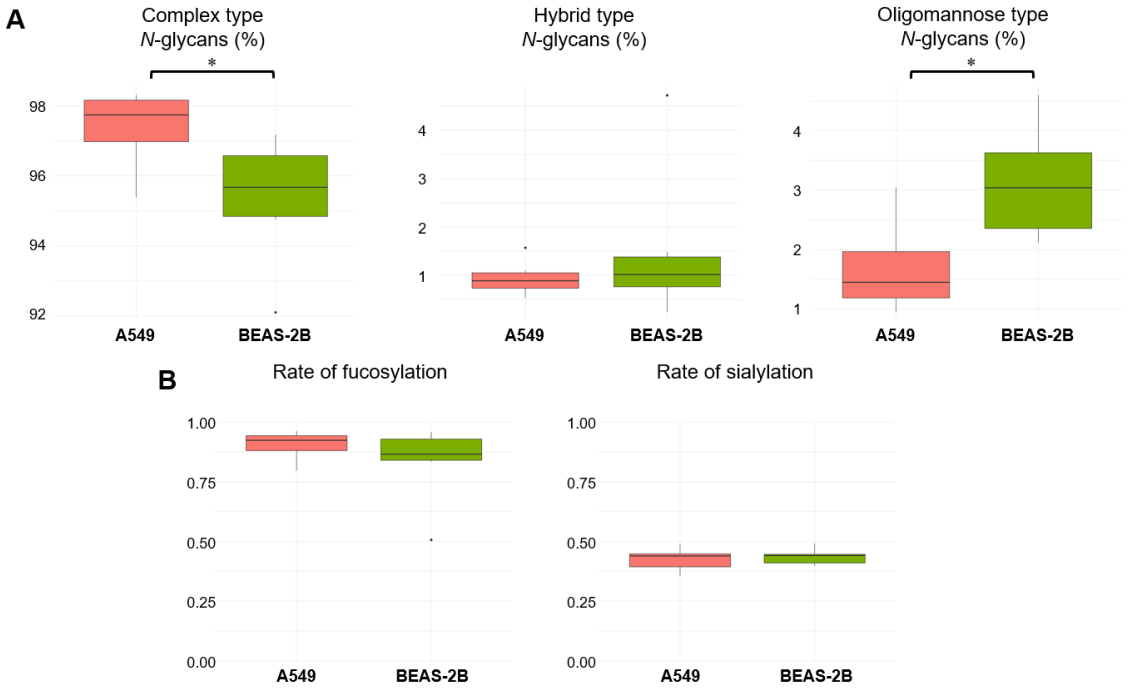

Figure 19. Relative abundance of *N*-glycan classes (**A**) and rates of fucosylation and sialylation (**B**). (*$p < 0.05$)

In both sEV types, complex *N*-glycans predominated, accounting for an average of 96.4% of all detected structures, while oligomannose and hybrid *N*-glycans accounted for 2.4% and 1.2%, respectively. The comparison of the sEV populations revealed minor differences. In A549 sEVs, the proportion of complex-type *N*-glycans displayed a slight but significant increase (FC = 1.02). This was accompanied by a significant decrease in oligomannose-type *N*-glycans (FC = 0.54) and a non-significant decrease in hybrid-type structures (FC = 0.62). Fucosylation showed a slight increase in A549 sEVs (FC = 1.14), while sialylation showed a slight decrease (FC = 0.97), however, these changes were statistically not significant.

4.2.4. CS/DS glycosaminoglycan analysis

In the CS/DS disaccharide analysis, four CS/DS disaccharides were analyzed across the 12 sEV samples. However, one A549 sEV sample did not contain detectable amounts of disaccharides, most likely due to a sample preparation issue, and was therefore excluded from further analysis. No CS/DS disaccharides were detected in either F12 or BEGM control media, confirming that the observed signals originated from the sEV samples rather than background contamination.

Overall, A549 sEVs had significantly higher CS/DS amounts, with total disaccharide amount showing an average 3.4-fold increase compared to BEAS-2B sEVs (**Fig. 20/A**). Analysis of the relative disaccharide composition revealed that the proportion of D0a6 was significantly reduced in A549 sEVs (FC = 0.76), while D0a4 showed a significant increase (FC = 1.59). In contrast, no significant changes were observed for D0a0 (FC = 0.65) or D0a10 (FC = 1.23). Boxplots illustrating the differences in relative abundances are presented in **Fig. 20/B**.

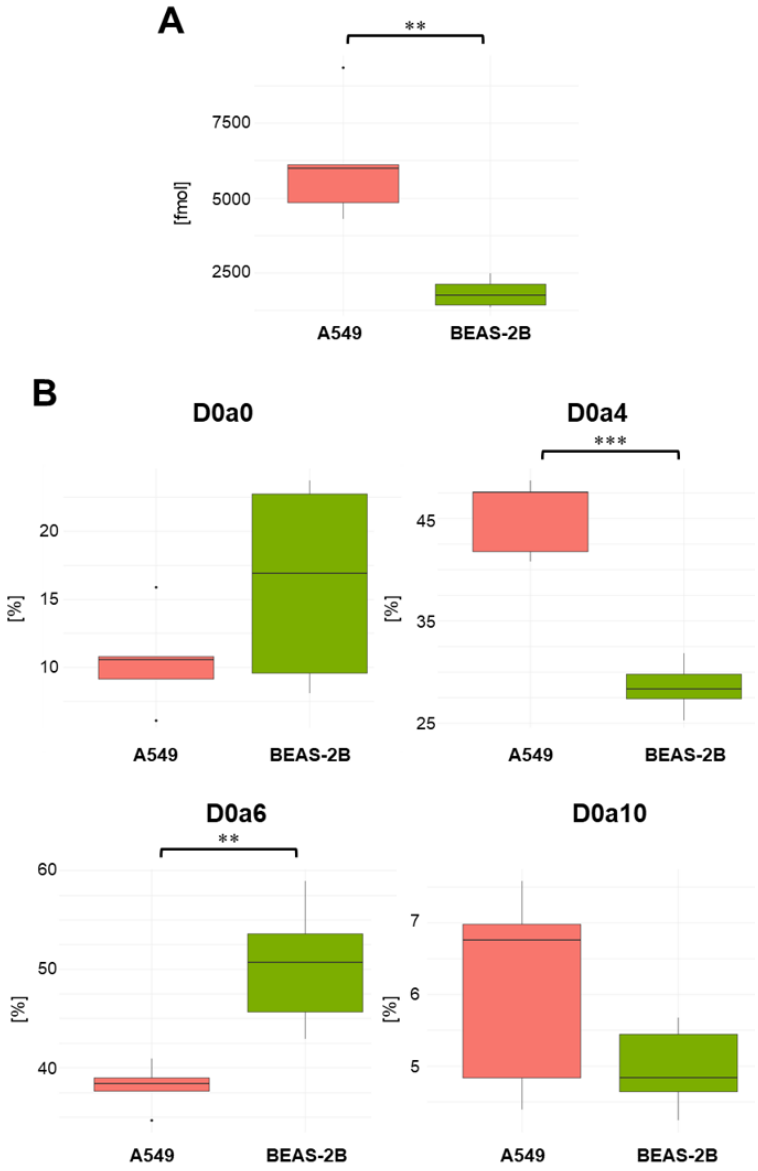

Figure 20. Total amount (**A**) and relative abundance (**B**) of CS/DS disaccharides in A549 and BEAS-2B sEVs. (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$)

These compositional differences resulted in a slight, non-significant increase in the overall degree of CS/DS sulfation (FC = 1.08; **Fig. 21/A**). At the same time, the 6S/4S ratio was significantly reduced in A549 sEVs (FC = 0.48; **Fig. 21/B**).

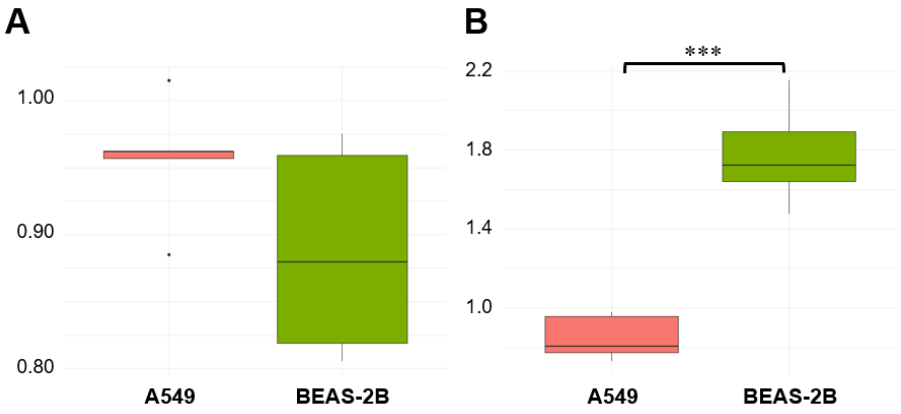

Figure 21. Average rate of CS/DS sulfation (**A**) and ratio of 6S/4S sulfated CS/DS disaccharides (**B**) in A549 and BEAS-2B sEVs. (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$)

PCA (**Fig. 22/A**) and clustered heatmap analyses (**Fig. 22/B**) were applied to characterize differences between the relative amounts of CS/DS disaccharides. Both methods were able to separate A549 and BEAS-2B sEVs completely, supporting the significant differences in CS/DS profiles between the two sEV groups. In PCA, PC1 and PC2 explained 63.2% and 29.0% of the variation.

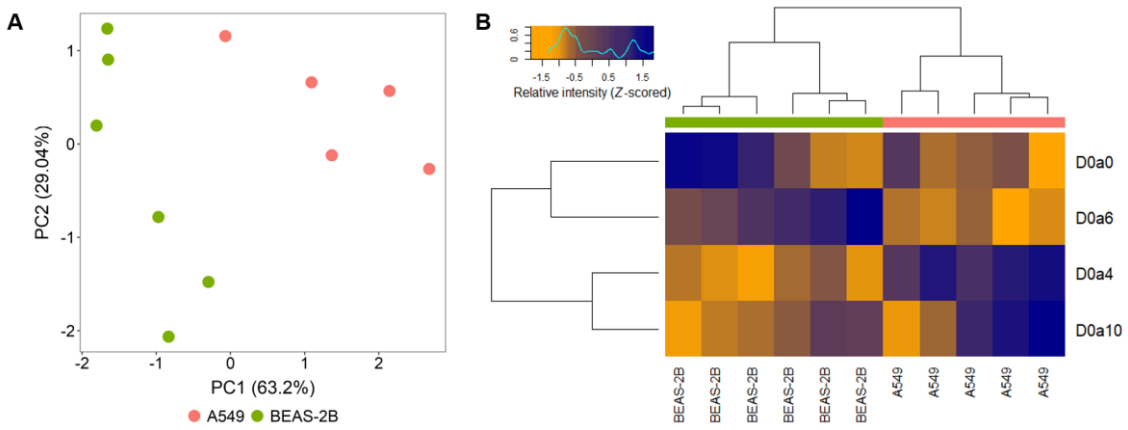

Figure 22. CS/DS principal component analysis (**A**) and clustered heatmap (**B**) on relative CS/DS abundances.

5. Discussion

5.1. Tissue study

The tissue study provided a comprehensive overview of protein expression and the composition of CS/DS and HS GAGs in morphologically distinct regions of lung tumor tissue, highlighting the molecular heterogeneity of the tumor microenvironment. By integrating proteomic and GAG disaccharide analyses, we were able to identify both protein-level differences and changes in glycosylation characteristics, which are closely related to ECM structure and cell signaling.

5.1.1. Proteomic analysis

An important finding of the proteomic analysis was the clear distinction between the ANL and the tumor regions, as confirmed by clustered heatmap, PCA, and the high number (250-541) of differentially expressed proteins. This suggests that tumor-related proteomic changes are pronounced despite intratumoral heterogeneity and are consistent across patients. At the same time, differences between tumor subtypes were also obvious, especially in the case of solid and papillary morphologies, between which 199 differentially expressed proteins were identified.

Since PGs are the carriers of GAG chains, particular attention was paid to PG core proteins. Among these, the expression of agrin and perlecan was consistently reduced in tumor regions compared to ANL. Both proteins are key components of the basement membrane and play a role in maintaining tissue structure as well as in signal transduction (126, 127). A decrease in their expression may reflect degradation or remodeling of the basement membrane, which facilitates tumor invasion. Similarly, the expression of collagen alpha-1(VIII) chain was reduced in solid tumors. No significant differences in PG core protein expression were observed among tumor subtypes, suggesting that while changes in the ECM are a general feature of tumorigenesis, tumor types differ less at the level of PG core proteins.

STRING functional enrichment analysis revealed that differentially expressed proteins between tumor subtypes were primarily associated with protein synthesis, including translation and RNA splicing. This suggests that alterations in RNA processing and protein synthesis represent an important distinguishing feature between morphological tumor types. Many of these proteins belonged to the class of RNA-binding proteins, which influence gene expression by regulating RNA processing, stability, and translation

(128). Dysregulation of these proteins has been implicated in cancer development through their effects on oncogenes and tumor suppressors (129, 130).

Differentially expressed proteins between ANL and tumor regions were associated not only with protein synthesis, but also with a broader range of altered biological processes (ECM organization, protein folding) and localizations (mitochondrion). The ECM is a dynamic and highly organized network composed of fibrous proteins and PGs, which determines the tissue structure and function (131). In cancer, the balance between ECM synthesis and degradation is disrupted, leading to extensive remodeling (132, 133). The decrease in the amount of ECM proteins observed in tumor areas is consistent with previous results in NSCLC tissues (134).

Increased abundances of proteins involved in protein folding indicate stress responses associated with the tumor. In the endoplasmic reticulum, protein homeostasis is maintained by quality control systems such as the unfolded protein response, which is activated when misfolded proteins accumulate (135, 136). Permanent activation of this pathway is a common feature of cancer and promotes the adaptation of tumor cells to increased biosynthesis and environmental stress (137).

Furthermore, the enrichment of mitochondrial proteins reflects alterations in cellular metabolism. Mitochondrial dynamics are regulated by a balance between fusion and fission; however, in cancer cells, fission often predominates (138, 139). This shift contributes to metabolic reprogramming and supports tumor cell proliferation and survival.

5.1.2. Glycosaminoglycan analysis

In PCA and clustered heatmap analyses, ANL and tumor regions formed separate clusters, indicating substantial differences in their GAG profiles. In contrast, tumor samples were not separated based on their morphological subtypes.

GAG analysis revealed significant alterations in both CS/DS and HS disaccharide profiles across the investigated tissue regions. Tumor regions showed a marked increase in total GAG amounts compared to ANL, with additional differences observed between tumor subtypes, particularly in the case of HS. While total GAG amounts had relatively low variability in ANL samples, considerable heterogeneity was observed in the tumor regions, suggesting that total GAG amounts alone are not sufficient to distinguish tumor

subtypes. The overall increase in GAG abundance is consistent with previous reports describing the accumulation of GAGs in tumor tissues (76, 77, 140).

Interestingly, these changes were not accompanied by increases in PG core protein abundances. In fact, the expression of several PG core proteins was reduced in tumor regions. This suggests that the differences observed in the amounts of GAG disaccharides were not caused by changes in core protein abundance, but rather by modifications at the level of GAG biosynthesis, such as increases in the number and length of the chains.

In addition to total GAG amounts, significant changes were observed in sulfation patterns, which have also been reported in other types of cancer, e.g., prostate (141), breast (142), and ovarian cancer (143). For CS/DS disaccharides, the overall rate of sulfation was higher in tumor regions, and a shift was observed in the relative distribution of 4S and 6S sulfated forms. The dominance of 6S sulfation in tubular and solid tumors, in contrast to the slight dominance of 4S sulfation observed in ANL and papillary regions, may indicate subtype-specific regulation of sulfotransferase activity. These observations are consistent with previous reports describing increased 6S sulfation accompanied by reduced 4S sulfation, structural changes that are thought to contribute to tumor microenvironment remodeling and to influence signal transduction, thereby promoting tumor development and progression (144).

For HS disaccharides, tumor regions showed increased sulfation rates. Interestingly, this contrasts with several reports describing decreased HS sulfation in cancer, highlighting the need for further investigation (111, 141, 145). At the same time, the relative contribution of *N*-sulfation decreased, while that of *O*-sulfation increased. This difference resulted primarily from changes in monosulfated and, to a lesser extent, in disulfated HS disaccharides, suggesting a shift in the balance of enzymatic modifications during HS biosynthesis. Altered expression of *N*-deacetylase/*N*-sulfotransferase enzymes has been reported in cancer and may contribute to the observed changes in the *N*/*O*-sulfation ratio (146, 147).

Finally, clustering results indicated a strong association between mucin content and GAG composition, as regions with moderate (M2) and high (M3) mucin content formed clearly distinct groups. This may reflect coordinated regulation of glycosyltransferases involved in mucin-type *O*-glycosylation and GAG biosynthesis. In contrast, stromal content showed only a minor influence on the overall GAG profiles.

5.2. Extracellular vesicle study

The EV study focused on the molecular characterization of sEVs released by lung cancer and non-tumorigenic epithelial cells. By combining proteomics, *N*-glycoproteomics, and GAG analyses, we aimed to capture both compositional and glycosylation-related differences that may reflect tumor-associated processes and intercellular communication through sEVs.

5.2.1. Characterization of small extracellular vesicles

The applied isolation and characterization workflow confirmed that the investigated particles corresponded to sEVs. Both TEM and MRPS analyses demonstrated that the vesicles fall within the expected size range (<200 nm), with A549 sEVs generally smaller and more homogeneous than BEAS-2B sEVs. The higher particle concentration observed in A549 samples suggests enhanced sEV production by cancer cells, which is consistent with previous observations that tumor cells show increased EV secretion (148, 149). This elevated sEV release may support tumor progression by facilitating communication with surrounding cells and modulating the tumor microenvironment.

5.2.2. Proteomic analysis

Clear differences were observed in the protein composition of A549 and BEAS-2B sEVs. PCA and clustered heatmap showed complete separation, indicating that the origin of the sEVs strongly determines their proteomic profiles.

The presence of a large proportion of proteins listed in the ExoCarta Top 100 database confirmed the reliability of the sEV isolation and the biological relevance of the dataset. However, we also observed considerable background protein contamination in the media, which can significantly distort the apparent proteomic profiles. This underlines the importance of strict selection criteria.

Of the 945 proteins quantified, 408 proteins were differentially expressed between A549 and BEAS-2B sEVs, indicating extensive remodeling of the sEV cargo. Several of these proteins have previously been reported to be dysregulated in cancer, such as fibulin-2, an ECM protein involved in cell adhesion and tissue organization (150). In a lung cancer cell line study, fibulin-2 was underexpressed in 9 of 11 cell lines compared to normal bronchial epithelial cells (151). In contrast, fibronectin and proliferating cell nuclear antigen were overexpressed in A549 sEVs. Increased amount of fibronectin has been described in plasma EVs of breast cancer patients (152), while proliferating cell nuclear

antigen was upregulated in lung cancer tissue (153). These findings support the concept that the molecular cargo of EVs reflects the physiological and pathological state of their cells of origin, as the observed alterations are consistent with cancer-associated changes at the cellular level (99, 154).

Several CSPG core proteins showed altered expression patterns, including both up- and downregulations. Cell surface PGs (CSPG4, syndecan-4) were downregulated in A549 sEVs, while ECM PGs (versican, aggrecan, and testican-1) showed mixed alterations (155). The abundance of testican-1 increased 14.3-fold in A549 sEVs. This protein is overexpressed in various types of cancer, including lung cancer, and its presence has also been confirmed in EVs (156, 157). On the other hand, syndecan-4 was 12.0 times more abundant in BEAS-2B sEVs. This protein plays a key role in cell adhesion and signal transduction, and its amount is elevated in some types of cancer and reduced in others (158). In addition, it can undergo shedding, generating soluble forms with distinct biological functions (159).

GSEA highlighted functional differences between the two sEV populations. Enrichment of pathways related to cell cycle regulation, DNA repair, metabolism, and protein biosynthetic processes in A549 sEVs reflects the proliferative and metabolically active phenotype of tumor cells. At the same time, the reduced representation of immune response processes suggests that tumor sEVs may contribute to immune evasion. These functional patterns are consistent with known cancer-associated processes and further support the role of tumor EVs as carriers of the molecular characteristics of their parental cells (160, 161).

5.2.3. Glycomics guided *N*-glycoproteomic analysis

Studies addressing *N*-glycosylation in EVs remain relatively scarce and most available data originates from urine and blood (162). An earlier study comparing sEVs from SCLC and NSCLC cell lines demonstrated that NSCLC sEVs are enriched in core-fucosylated bi- and triantennary *N*-glycans, while SCLC sEVs display greater *N*-glycan heterogeneity (103).

In this study, the integration of *N*-glycomics and *N*-glycoproteomics enabled a detailed, site-specific characterization of *N*-glycosylation in sEVs. Initial *N*-glycan profiling confirmed that the vast majority of the identified *N*-glycans originated from the sEVs

rather than from the culture media, supporting the validity of subsequent *N*-glycopeptide analysis.

PCA and clustered heatmap demonstrated a clear separation between A549 and BEAS-2B sEVs based on their *N*-glycopeptide profiles. This distinction was further supported by the high proportion of differentially represented *N*-glycoforms (152 out of 227). These differences are likely related to altered activity of glycosyltransferases and glycosidases, which are often dysregulated in cancer and can be linked to disease progression (163, 164).

Several highly *N*-glycosylated ECM proteins, including versican, laminins, and galectin-3-binding protein (LG3BP), showed significant changes at the *N*-glycoform level. Versican, a PG core protein involved in cell adhesion and migration (165), displayed site-specific alterations, with decreased abundance of complex, fucosylated multi-antennary *N*-glycans at one site (N1898) and increased representation of fucosylated biantennary structures at other positions. Laminin subunits, which play an important role in basement membrane organization (166), displayed primarily increased abundance of various *N*-glycoforms. Galectin-3-binding protein, a heavily *N*-glycosylated protein commonly found in EVs (167), exhibited widespread alterations, with most of its *N*-glycoforms showing reduced abundance in A549 sEVs. Notably, all three protein families have been previously implicated in cancer-related processes (168-171), suggesting that the observed *N*-glycosylation changes may contribute to their functional roles in tumor progression.

N-glycopeptide FCs were normalized to corresponding protein abundances to distinguish between changes driven by protein abundance and those reflecting changes in the degree of *N*-glycosylation. This approach revealed that a large proportion of the observed differences arose from alterations in *N*-glycan abundance, highlighting the importance of site-specific analysis.

When *N*-glycosylation profiles were analyzed at the *N*-glycan level, a slight increase in complex *N*-glycans and a decrease in oligomannose structures were observed in A549 sEVs. This difference is consistent with cancer-associated *N*-glycosylation patterns, which are characterized by increased amounts of complex, mostly highly branched *N*-glycans and enhanced glycosyltransferase activity (31, 163).

A slight, statistically non-significant increase was observed in fucosylation in A549 sEVs, which may reflect increased activity of fucosyltransferases such as FUT8, which is often upregulated in cancer (172). In contrast, sialylation showed only a minor difference between the two sEV populations. However, in this study, Neu5Ac linkage-specific information was not assessed, although previous reports highlighted the importance of α2,3- and α2,6-linked Neu5Ac in cancer (173). Therefore, exploring potential differences at the level of Neu5Ac linkage would be of interest in future studies.

5.2.4. CS/DS glycosaminoglycan analysis

Research on the GAG analysis of EVs is extremely scarce. To the best of our knowledge, this study is the first to use MS-based disaccharide-level analysis to characterize CS/DS found in EVs. Previous studies relied on gas chromatography-MS analysis of building blocks and showed that EVs contain large amounts of CS, DS, and KS (174, 175).

No CS/DS disaccharides were detected in the control media, confirming that the measured signals originated from sEVs and not from background contamination. PCA and clustered heatmap clearly distinguished A549 and BEAS-2B sEVs based on their CS/DS profiles.

Quantitative analysis revealed a significant increase in total CS/DS amount in A549 sEVs, with an approximately 3.4-fold higher disaccharide abundance compared to BEAS-2B sEVs. This observation is consistent with previous reports describing elevated CS/DS amounts in tumor tissues across various cancer types, including lung cancer (77, 111, 141). Changes occurred in both directions in PG core proteins, thus the increase in total GAG disaccharide amounts can be partly explained by changes in PG abundance.

In terms of the relative amounts of individual disaccharides, a significant increase in D0a4 was observed, accompanied by a significant decrease in D0a6, while D0a0 and D0a10 remained largely unchanged. As a result, average rate of CS/DS sulfation showed a slight but statistically insignificant increase, while the 6S/4S ratio showed a significant decrease.

Tissue studies provide a useful framework for interpreting the results. In lung cancer tissues, increased total CS/DS amounts and alterations in sulfation patterns have been consistently reported (76, 77, 104). Although the overall increase in CS/DS observed in A549 sEVs is consistent with these reports, an increase in 6S sulfation and a decrease in 4S sulfation were typically observed in tumor tissues (76, 104), while we observed the opposite in sEVs.

The decreased 6S/4S ratio may indicate changes in the activity of specific sulfotransferases, such as carbohydrate sulfotransferase 11, which is a chondroitin 4-*O*-sulfotransferase; its overexpression has been linked to a poor prognosis in cancer (176, 177).

Furthermore, the increased 4S ratio may be associated with the presence of oncofetal CS (ofCS). OfCS is a specific GAG structure that is normally present only in the placenta and is characterized by increased 4S sulfation (178, 179). Using VAR2CSA antibody, ofCS has been detected in a wide range of cancer cell lines and tumor tissues, indicating that this fetal motif is reactivated in cancer (180, 181). Although ofCS was not specifically examined in this study, our results suggested that tumor EVs may also carry this pattern, a finding that has since been confirmed by others (182, 183).

5.3. Limitations of the studies

The studies presented in this thesis have several limitations that should be considered when interpreting the results. First, both projects were based on relatively small sample sizes. The applied proteomic, *N*-glycoproteomic, and GAG analysis workflows are primarily discovery-oriented and provide descriptive molecular profiles rather than direct mechanistic insights. Therefore, the observed molecular differences and enriched biological processes should be interpreted as hypotheses-generating findings that require further functional validation.

In the tissue-based study, the number of biological replicates within individual morphological groups was limited, and some regions originated from the same patient, reducing full sample independence and potentially influencing statistical analyses and clustering results. Although spatially targeted surface digestion preserved regional information, it did not provide the single-cell or pixel-level spatial resolution comparable to imaging-based technologies.

In the extracellular vesicle study, EVs were isolated from immortalized cell lines (A549 and BEAS-2B), which cannot fully reproduce the complexity and heterogeneity of circulating extracellular vesicle populations in vivo. In addition, complete separation of EV subpopulations and non-vesicular contaminants cannot be achieved. While site-specific *N*-glycosylation changes were identified, detailed structural characterization and functional roles of individual *N*-glycoforms were beyond the scope of this work.

Functional validation of these findings would require targeted perturbation approaches, such as genetic or enzymatic modulation of glycosylation pathways.

Despite these limitations, the established workflows and datasets provide comprehensive molecular characterization of lung cancer tissues and extracellular vesicles, forming a strong foundation for future mechanistic and translational studies.

6. Conclusions

6.1. Tissue study

The results of the tissue study demonstrated that both proteomic composition and GAG profiles are substantially altered in lung cancer tissues compared to adjacent normal regions. Proteomic analysis showed clear differences between tumor and non-tumor regions, reflecting changes in biological processes related to extracellular matrix organization, protein folding, and protein synthesis. Less differences were observed between morphological subtypes, and differentially expressed proteins were associated mainly with protein synthesis.

GAG analysis showed a pronounced increase in total CS/DS and HS disaccharide amounts in tumor regions, accompanied by significant alterations in sulfation patterns. These changes were not directly associated with proteoglycan core protein abundance. The observed shifts in CS/DS 4S and 6S sulfation, as well as in HS *N*- and *O*-sulfation, suggest altered enzymatic regulation.

Overall, the combined proteomic and GAG analysis provided complementary insights into tumor heterogeneity and revealed molecular alterations that are not accessible through conventional bulk tissue approaches.

6.2. Extracellular vesicle study

The EV study demonstrated that sEVs derived from lung cancer and non-tumorigenic epithelial cells have distinct proteomic, *N*-glycoproteomic, and CS/DS GAG signatures. Proteomic analysis showed extensive remodeling of the sEV cargo in A549 sEVs, with a high number of differentially expressed proteins. These differences were associated with biological processes linked to proliferation, metabolism, and immune-related functions.

The integration of *N*-glycomics and *N*-glycoproteomics enabled confident identification of *N*-glycopeptides and provided site-specific insights into glycosylation changes. Versican, laminins, and galectin-3-binding protein possessed several dysregulated *N*-glycoforms and were previously implicated in cancer.

In addition, CS/DS analysis revealed a significant increase in total GAG amount in A549 sEVs, along with a shift toward increased 4S and decreased 6S sulfation. This pattern differs from that typically observed in tumor tissues, suggesting that GAG composition in sEVs is selectively regulated. The presence of ofCS further supports the biological relevance of these findings.

Together, these results demonstrate that sEV glycosylation provides valuable insight into the molecular mechanisms of lung cancer and represents a potential basis for the development of novel biomarker strategies.

6.3. Outlook

The findings of the second study highlight the strong potential of EVs as a non-invasive source of molecular biomarkers. Building on these results, further work has been initiated to better understand the relationship between cellular origin and sEV composition. In this ongoing study, multiple lung-derived cell lines representing different lung cancer subtypes (AC, SqCC, and LCC), as well as non-tumorigenic epithelial cells, are being investigated. In addition to sEVs, whole-cell lysates and isolated cell membrane fractions are also analyzed using proteomic, *N*-glycoproteomic, and GAG profiling approaches. This strategy enables direct comparison between cellular compartments and their corresponding sEVs, providing insight into the mechanisms governing sEV cargo selection.

In addition to cell line-based systems, we have also initiated research aimed at applying these methods to clinically relevant samples. Small EVs are isolated from human plasma, and their comprehensive molecular characterization is then performed. After method optimization for plasma, the goal is to apply these workflows to larger cohorts of lung cancer patients to identify *N*-glycoproteins, *N*-glycopeptides, and GAG characteristics that are specific to lung cancer, its subtypes, or the patient's therapeutic response.

Taken together, these future directions aim to bridge the gap between fundamental molecular characterization and clinical application, with the long-term objective of developing EV-based, minimally invasive biomarkers for lung cancer diagnosis and patient stratification.

7. Summary

Lung cancer is characterized by complex molecular alterations that go beyond changes in protein expression and also include alterations in glycosylation. Among these, *N*-glycosylation and GAGs play essential roles in cell signaling, extracellular matrix organization, and tumor progression. This thesis aimed to investigate these molecular features using advanced HPLC-MS-based analytical approaches, combining proteomics with in-depth glycosylation analysis in two complementary biological systems.

The first part of the work focused on spatially resolved lung cancer tissues carrying ALK rearrangement. By applying on-surface digestion techniques, both protein composition as well as CS/DS and HS GAG disaccharide profiles were analyzed from morphologically distinct regions. The results revealed clear molecular differences between tumor and adjacent non-tumor regions, particularly in pathways related to extracellular matrix organization and protein biosynthesis. In parallel, GAG analysis showed an increased overall abundance and altered sulfation patterns in tumor regions. Between tumor subtypes, the most prominent dysregulated processes were associated with protein synthesis, and the largest differences were observed in the total amount of HS disaccharides.

The second study investigated sEVs derived from A549 lung adenocarcinoma and BEAS-2B non-tumorigenic epithelial cell lines. Comprehensive characterization included proteomic, glycomics guided *N*-glycoproteomic, and CS/DS GAG analyses. The results demonstrated that all three molecular profiles clearly distinguished the two sEVs types. Differential expression of several PG core proteins was observed, and they showed mixed patterns. Several proteins, including laminins, galectin-3-binding protein, and versican, exhibited multiple *N*-glycosylation alterations. Additionally, GAG analysis revealed an increased total amount of CS disaccharides along with a shift toward higher 4S sulfation in A549 sEVs.

Overall, this work highlights the importance of combining proteomic and glycosylation approaches to achieve a more comprehensive understanding of lung cancer biology. These results contribute to the growing field of cancer glycobiology and support the potential application of glycosylation patterns in future biomarker discovery and disease characterization.

8. References

1. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev. 1996;13:19-50.
2. Zhong Q, Xiao X, Qiu Y, Xu Z, Chen C, Chong B, et al. Protein posttranslational modifications in health and diseases: Functions, regulatory mechanisms, and therapeutic implications. MedComm (2020). 2023;4(3):e261.
3. Cho WC. Proteomics technologies and challenges. Genomics Proteomics Bioinformatics. 2007;5(2):77-85.
4. Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet. 2013;14(1):35-48.
5. Kwon YW, Jo HS, Bae S, Seo Y, Song P, Song M, et al. Application of Proteomics in Cancer: Recent Trends and Approaches for Biomarkers Discovery. Front Med (Lausanne). 2021;8:747333.
6. Gheybi E, Hosseinzadeh P, Tayebi-Khorrami V, Rostami M, Soukhtanloo M. Proteomics in decoding cancer: A review. Clin Chim Acta. 2025;574:120302.
7. Yang XL, Shi Y, Zhang DD, Xin R, Deng J, Wu TM, et al. Quantitative proteomics characterization of cancer biomarkers and treatment. Mol Ther Oncolytics. 2021;21:255-63.
8. Dupree EJ, Jayathirtha M, Yorkey H, Mihasan M, Petre BA, Darie CC. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of this Field. Proteomes. 2020;8(3).
9. Hovey OFJ, Lajoie GA, Cooper TT. Middle-down proteomics: the pursuit for longer peptides. Expert Rev Proteomics. 2025;22(11-12):453-70.
10. Nickerson JL, Baghalabadi V, Rajendran S, Jakubec PJ, Said H, McMillen TS, et al. Recent advances in top-down proteome sample processing ahead of MS analysis. Mass Spectrom Rev. 2023;42(2):457-95.
11. Miller RM, Smith LM. Overview and considerations in bottom-up proteomics. Analyst. 2023;148(3):475-86.
12. Alanazi S. Recent Advances in Liquid Chromatography-Mass Spectrometry (LC-MS) Applications in Biological and Applied Sciences. Anal Sci Adv. 2025;6(1):e70024.

13. Nadler WM, Waidelich D, Kerner A, Hanke S, Berg R, Trumpp A, et al. MALDI versus ESI: The Impact of the Ion Source on Peptide Identification. J Proteome Res. 2017;16(3):1207-15.
14. Wilson SR, Vehus T, Berg HS, Lundanes E. Nano-LC in proteomics: recent advances and approaches. Bioanalysis. 2015;7(14):1799-815.
15. Shen Y, Zhao R, Berger SJ, Anderson GA, Rodriguez N, Smith RD. High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. Anal Chem. 2002;74(16):4235-49.
16. Bian Y, Zheng R, Bayer FP, Wong C, Chang YC, Meng C, et al. Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. Nat Commun. 2020;11(1):157.
17. Haag AM. Mass Analyzers and Mass Spectrometers. Adv Exp Med Biol. 2016;919:157-69.
18. Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. Mol Cell Proteomics. 2018;17(12):2534-45.
19. Lancaster NM, Sinitcyn P, Forny P, Peters-Clarke TM, Fecher C, Smith AJ, et al. Fast and deep phosphoproteome analysis with the Orbitrap Astral mass spectrometer. Nat Commun. 2024;15(1):7016.
20. Neagu AN, Jayathirtha M, Baxter E, Donnelly M, Petre BA, Darie CC. Applications of Tandem Mass Spectrometry (MS/MS) in Protein Analysis for Biomedical Research. Molecules. 2022;27(8).
21. Coon JJ, Syka JEP, Shabanowitz J, Hunt DF. Tandem Mass Spectrometry for Peptide and Protein Sequence Analysis. BioTechniques. 2005;38(4):519-23.
22. Bateman NW, Goulding SP, Shulman NJ, Gadok AK, Szumlinski KK, MacCoss MJ, et al. Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). Mol Cell Proteomics. 2014;13(1):329-38.
23. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol. 2018;14(8):e8126.
24. Cottrell JS. Protein identification using MS/MS data. J Proteomics. 2011;74(10):1842-51.

25. Rozanova S, Barkovits K, Nikolov M, Schmidt C, Urlaub H, Marcus K. Quantitative Mass Spectrometry-Based Proteomics: An Overview. Methods Mol Biol. 2021;2228:85-116.
26. Patel VJ, Thalassinos K, Slade SE, Connolly JB, Crombie A, Murrell JC, et al. A comparison of labeling and label-free mass spectrometry-based proteomics approaches. J Proteome Res. 2009;8(7):3752-9.
27. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.
28. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer. 2015;15(9):540-55.
29. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol. 2012;13(7):448-62.
30. Peixoto A, Relvas-Santos M, Azevedo R, Santos LL, Ferreira JA. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. Front Oncol. 2019;9:380.
31. Lin Y, Lubman DM. The role of N-glycosylation in cancer. Acta Pharm Sin B. 2024;14(3):1098-110.
32. Xu X, Peng Q, Jiang X, Tan S, Yang W, Han Y, et al. Altered glycosylation in cancer: molecular functions and therapeutic potential. Cancer Commun (Lond). 2024;44(11):1316-36.
33. Duarte HO, Reis CA, Blanchard V, Tauber R. Glycosylation in Cancer. Handb Exp Pharmacol. 2025;288:243-93.
34. Balbisi M, Sugár S, Turiák L. Protein glycosylation in lung cancer from a mass spectrometry perspective. Mass Spectrom Rev. 2026;45(3):455-75.
35. Kirwan A, Utratna M, O'Dwyer ME, Joshi L, Kilcoyne M. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. Biomed Res Int. 2015;2015:490531.
36. Tkac J, Gajdosova V, Hroncekova S, Bertok T, Hires M, Jane E, et al. Prostate-specific antigen glycoprofiling as diagnostic and prognostic biomarker of prostate cancer. Interface Focus. 2019;9(2):20180077.

37. Zhou JM, Wang T, Zhang KH. AFP-L3 for the diagnosis of early hepatocellular carcinoma: A meta-analysis. Medicine (Baltimore). 2021;100(43):e27673.
38. Stanley P, Moremen KW, Lewis NE, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022. p. 103-16.
39. Yang Y, Franc V, Heck AJR. Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis. Trends Biotechnol. 2017;35(7):598-609.
40. Rafique S, Yang S, Sajid MS, Faheem M. A review of intact glycopeptide enrichment and glycan separation through hydrophilic interaction liquid chromatography stationary phase materials. J Chromatogr A. 2024;1735:465318.
41. Sun Z, Lih TM, Woo J, Jiao L, Hu Y, Wang Y, et al. Improving Glycoproteomic Analysis Workflow by Systematic Evaluation of Glycopeptide Enrichment, Quantification, Mass Spectrometry Approach, and Data Analysis Strategies. Anal Chem. 2024;96(52):20481-90.
42. Riley NM, Malaker SA, Driessen MD, Bertozzi CR. Optimal Dissociation Methods Differ for N- and O-Glycopeptides. J Proteome Res. 2020;19(8):3286-301.
43. Sandilya V, Sahioun S, Suma TA, Adewolu A, Koraich A, Nishe SS, et al. Advances in Fragmentation Techniques for Glycomics and Glycoproteomics. Mass Spectrom Rev. 2025.
44. Chang D, Zaia J. Methods to improve quantitative glycoprotein coverage from bottom-up LC-MS data. Mass Spectrom Rev. 2022;41(6):922-37.
45. Jager S, Zeller M, Pashkova A, Schulte D, Damoc E, Reiding KR, et al. In-depth plasma N-glycoproteome profiling using narrow-window data-independent acquisition on the Orbitrap Astral mass spectrometer. Nat Commun. 2025;16(1):2497.
46. Cao W. Advancing mass spectrometry-based glycoproteomic software tools for comprehensive site-specific glycoproteome analysis. Curr Opin Chem Biol. 2024;80:102442.
47. Reid DJ, Thibert S, Zhou M. Dissecting the structural heterogeneity of proteins by native mass spectrometry. Protein Sci. 2023;32(4):e4612.

48. Merry CLR, Lindahl U, Couchman J, Esko JD. Proteoglycans and Sulfated Glycosaminoglycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022. p. 217-32.
49. Casale J, Crane JS. Biochemistry, Glycosaminoglycans. Treasure Island (FL): StatPearls Publishing; 2026.
50. Zappe A, Miller RL, Struwe WB, Pagel K. State-of-the-art glycosaminoglycan characterization. Mass Spectrom Rev. 2022;41(6):1040-71.
51. Lawrence R, Lu H, Rosenberg RD, Esko JD, Zhang L. Disaccharide structure code for the easy representation of constituent oligosaccharides from glycosaminoglycans. Nat Methods. 2008;5(4):291-2.
52. Pepi LE, Sanderson P, Stickney M, Amster IJ. Developments in Mass Spectrometry for Glycosaminoglycan Analysis: A Review. Mol Cell Proteomics. 2021;20:100025.
53. Perez S, Makshakova O, Angulo J, Bedini E, Bisio A, de Paz JL, et al. Glycosaminoglycans: What Remains To Be Deciphered? JACS Au. 2023;3(3):628-56.
54. Tóth G, Vékey K, Drahos L, Horváth V, Turiák L. Salt and solvent effects in the microscale chromatographic separation of heparan sulfate disaccharides. J Chromatogr A. 2020;1610:460548.
55. Tóth G, Vékey K, Sugár S, Kovalszky I, Drahos L, Turiák L. Salt gradient chromatographic separation of chondroitin sulfate disaccharides. J Chromatogr A. 2020;1619:460979.
56. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-63.
57. Luo G, Zhang Y, Rumgay H, Morgan E, Langselius O, Vignat J, et al. Estimated worldwide variation and trends in incidence of lung cancer by histological subtype in 2022 and over time: a population-based study. Lancet Respir Med. 2025;13(4):348-63.
58. Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. Nat Rev Dis Primers. 2021;7(1):3.

59. Raman V, Yang CJ, Deng JZ, D'Amico TA. Surgical treatment for early stage non-small cell lung cancer. J Thorac Dis. 2018;10(Suppl 7):S898-s904.
60. Wang M, Herbst RS, Boshoff C. Toward personalized treatment approaches for non-small-cell lung cancer. Nat Med. 2021;27(8):1345-56.
61. Huang H. Anaplastic Lymphoma Kinase (ALK) Receptor Tyrosine Kinase: A Catalytic Receptor with Many Faces. Int J Mol Sci. 2018;19(11).
62. Lei Y, Lei Y, Shi X, Wang J. EML4-ALK fusion gene in non-small cell lung cancer. Oncol Lett. 2022;24(2):277.
63. Winter M, Jeroch J, Wetz M, Rauschendorf MA, Wild PJ. Molecular Pathology of Advanced NSCLC: Biomarkers and Therapeutic Decisions. Cancers (Basel). 2026;18(2).
64. Rulten SL, Grose RP, Gatz SA, Jones JL, Cameron AJM. The Future of Precision Oncology. Int J Mol Sci. 2023;24(16).
65. Thomas A, Rajan A, Giaccone G. Tyrosine kinase inhibitors in lung cancer. Hematol Oncol Clin North Am. 2012;26(3):589-605, viii.
66. Attili I, Corvaja C, Spitaleri G, Del Signore E, Trillo Aliaga P, Passaro A, et al. New Generations of Tyrosine Kinase Inhibitors in Treating NSCLC with Oncogene Addiction: Strengths and Limitations. Cancers (Basel). 2023;15(20).
67. Mohiuddin M. Anti-PD-1/PD-L1 Immunotherapy as a Potential Treatment Option for Lung Cancer: A Perspective Analysis of Opportunities and Challenges. Health Sci Rep. 2026;9(1):e71749.
68. Ou X, Gao G, Habaz IA, Wang Y. Mechanisms of resistance to tyrosine kinase inhibitor-targeted therapy and overcoming strategies. MedComm (2020). 2024;5(9):e694.
69. Li L, Pu H, Zhang X, Guo X, Li G, Zhang M. Resistance to PD-1/PD-L1 immune checkpoint blockade in advanced non-small cell lung cancer. Crit Rev Oncol Hematol. 2025;209:104683.
70. Alvarez MR, Zhou Q, Tena J, Barboza M, Wong M, Xie Y, et al. Glycomic, Glycoproteomic, and Proteomic Profiling of Philippine Lung Cancer and Peritumoral Tissues: Case Series Study of Patients Stages I-III. Cancers (Basel). 2023;15(5).

71. Jezková P, Skřičková J, Wimmer G, Jr., Zelinková J, Zdráhal Z, Lattová E. Differentiation of Sialyl Linkages Using a Combination of Alkyl Esterification and Phenylhydrazine Derivatization: Application for N-Glycan Profiling in the Sera of Patients with Lung Cancer. Anal Chem. 2022;94(18):6736-44.
72. Alvarez MRS, Zhou Q, Tena J, Lebrilla CB, Completo GC, Heralde FM, 3rd, et al. N-Glycan and Glycopeptide Serum Biomarkers in Philippine Lung Cancer Patients Identified Using Liquid Chromatography-Tandem Mass Spectrometry. ACS Omega. 2022;7(44):40230-40.
73. Gao Z, Wu Z, Han Y, Zhang X, Hao P, Xu M, et al. Aberrant Fucosylation of Saliva Glycoprotein Defining Lung Adenocarcinomas Malignancy. ACS Omega. 2022;7(21):17894-906.
74. Soltermann A, Ossola R, Kilgus-Hawelski S, von Eckardstein A, Suter T, Aebersold R, et al. N-glycoprotein profiling of lung adenocarcinoma pleural effusions by shotgun proteomics. Cancer. 2008;114(2):124-33.
75. Li QK, Shah P, Li Y, Aiyetan PO, Chen J, Yung R, et al. Glycoproteomic analysis of bronchoalveolar lavage (BAL) fluid identifies tumor-associated glycoproteins from lung adenocarcinoma. J Proteome Res. 2013;12(8):3689-96.
76. Li G, Li L, Joo EJ, Son JW, Kim YJ, Kang JK, et al. Glycosaminoglycans and glycolipids as potential biomarkers in lung cancer. Glycoconj J. 2017;34(5):661-9.
77. Pál D, Tóth G, Sugár S, Fügedi KD, Szabó D, Kovalszky I, et al. Compositional Analysis of Glycosaminoglycans in Different Lung Cancer Types-A Pilot Study. Int J Mol Sci. 2023;24(8).
78. Pál D, Bugyi F, Virág D, Szabó D, Schlosser G, Kovalszky I, et al. Analysis and characterization of chondroitin/dermatan sulfate composition of lung adenocarcinoma tissues with different types of genetic alterations in ALK, EGFR and KRAS oncogenes. Carbohydrate Polymer Technologies and Applications. 2025;10:100826.
79. Zhao Y, Shen M, Wu L, Yang H, Yao Y, Yang Q, et al. Stromal cells in the tumor microenvironment: accomplices of tumor progression? Cell Death Dis. 2023;14(9):587.

80. Eguchi T, Kadota K, Park BJ, Travis WD, Jones DR, Adusumilli PS. The new IASLC-ATS-ERS lung adenocarcinoma classification: what the surgeon should know. Semin Thorac Cardiovasc Surg. 2014;26(3):210-22.
81. Wang Y, Hu J, Sun Y, Lu Y. Micropapillary or solid component predicts worse prognosis in pathological IA stage lung adenocarcinoma: A meta-analysis. Medicine (Baltimore). 2023;102(49):e36503.
82. Zha L, Matsu-Ura T, Sluka JP, Murakawa T, Tsuta K. Morphological basis of the lung adenocarcinoma subtypes. iScience. 2024;27(5):109742.
83. Wegrowski Y, Maquart FX. Involvement of stromal proteoglycans in tumour progression. Crit Rev Oncol Hematol. 2004;49(3):259-68.
84. Chen C, Patel A, Demirkhanyan L, Gondi CS. The Role of Mucins in Cancer and Cancer Progression: A Comprehensive Review. Curr Issues Mol Biol. 2025;47(6).
85. Hussaini HM, Seo B, Rich AM. Immunohistochemistry and Immunofluorescence. Methods Mol Biol. 2023;2588:439-50.
86. Uhlén M, Björling E, Agaton C, Szigyarto CA, Amini B, Andersen E, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics. 2005;4(12):1920-32.
87. Moore JL, Charkoftaki G. A Guide to MALDI Imaging Mass Spectrometry for Tissues. J Proteome Res. 2023;22(11):3401-17.
88. Mund A, Coscia F, Kriston A, Hollandi R, Kovács F, Brunner AD, et al. Deep Visual Proteomics defines single-cell identity and heterogeneity. Nat Biotechnol. 2022;40(8):1231-40.
89. Conroy LR, Chang JE, Sun Q, Clarke HA, Buoncristiani MD, Young LEA, et al. High-dimensionality reduction clustering of complex carbohydrates to study lung cancer metabolic heterogeneity. Adv Cancer Res. 2022;154:227-51.
90. Devlin A, Green F, Takats Z. Mass Spectrometry Imaging with Trapped Ion Mobility Spectrometry Enables Spatially Resolved Chondroitin, Dermatan, and Hyaluronan Glycosaminoglycan Oligosaccharide Analysis In Situ. Anal Chem. 2024;96(45):17969-77.
91. Palomino TV, Muddiman DC. Glycosaminoglycan Mass Spectrometry Imaging by Infrared Matrix-Assisted Laser Desorption Electrospray Ionization. J Am Soc Mass Spectrom. 2025;36(4):658-63.

92. Turiák L, Shao C, Meng L, Khatri K, Leymarie N, Wang Q, et al. Workflow for combined proteomics and glycomics profiling from histological tissues. Anal Chem. 2014;86(19):9670-8.
93. Sugár SN, Molnár BA, Bugyi F, Kecskeméti G, Szabó Z, Laczó I, et al. Glycoproteomics Analysis of Triple Wild-Type Lung Adenocarcinoma Tissue Samples. J Proteome Res. 2025;24(5):2419-29.
94. Maacha S, Bhat AA, Jimenez L, Raza A, Haris M, Uddin S, et al. Extracellular vesicles-mediated intercellular communication: roles in the tumor microenvironment and anti-cancer drug resistance. Mol Cancer. 2019;18(1):55.
95. Ragni E. Extracellular Vesicles: Recent Advances and Perspectives. Front Biosci (Landmark Ed). 2025;30(6):36405.
96. Kalluri R, McAndrews KM. The role of extracellular vesicles in cancer. Cell. 2023;186(8):1610-26.
97. Welsh JA, Goberdhan DCI, O'Driscoll L, Buzas EI, Blenkiron C, Bussolati B, et al. Minimal information for studies of extracellular vesicles (MISEV2023): From basic to advanced approaches. J Extracell Vesicles. 2024;13(2):e12404.
98. Wang Y, Zhao R, Jiao X, Wu L, Wei Y, Shi F, et al. Small Extracellular Vesicles: Functions and Potential Clinical Applications as Cancer Biomarkers. Life (Basel). 2021;11(10).
99. Kumar MA, Baba SK, Sadida HQ, Marzooqi SA, Jerobin J, Altemani FH, et al. Extracellular vesicles as tools and targets in therapy for diseases. Signal Transduct Target Ther. 2024;9(1):27.
100. An T, Qin S, Sun D, Huang Y, Hu Y, Li S, et al. Unique Protein Profiles of Extracellular Vesicles as Diagnostic Biomarkers for Early and Advanced Non-Small Cell Lung Cancer. Proteomics. 2019;19(12):e1800160.
101. Novikova SE, Soloveva NA, Farafonova TE, Tikhonova OV, Liao PC, Zgoda VG. Proteomic Signature of Extracellular Vesicles for Lung Cancer Recognition. Molecules. 2021;26(20).
102. Rosa-Fernandes L, Rocha VB, Carregari VC, Urbani A, Palmisano G. A Perspective on Extracellular Vesicles Proteomics. Front Chem. 2017;5:102.

103. Kondo K, Harada Y, Nakano M, Suzuki T, Fukushige T, Hanzawa K, et al. Identification of distinct N-glycosylation patterns on extracellular vesicles from small-cell and non-small-cell lung cancer cells. J Biol Chem. 2022;298(6):101950.
104. Balbisi M, Sugár S, Schlosser G, Szeitz B, Fillinger J, Moldvay J, et al. Inter- and intratumoral proteomics and glycosaminoglycan characterization of ALK rearranged lung adenocarcinoma tissues: a pilot study. Sci Rep. 2023;13(1):6268.
105. Balbisi M, Langó T, Horváth VN, Pál D, Schlosser G, Kecskeméti G, et al. Distinct Proteomic and Glycosylation Signatures Differentiate A549 Tumor and BEAS-2B Nontumor Cell Line-Derived Small Extracellular Vesicles. Mol Cell Proteomics. 2026;25(3):101524.
106. 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.
107. Tóth G, Sugár S, Balbisi M, Molnár BA, Bugyi F, Fügedi KD, et al. Optimized Sample Preparation and Microscale Separation Methods for High-Sensitivity Analysis of Hydrophilic Peptides. Molecules. 2022;27(19).
108. Bern M, Kil YJ, Becker C. Byonic: advanced peptide and protein identification software. Curr Protoc Bioinformatics. 2012;Chapter 13:13.20.1-13.20.14.
109. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016;11(12):2301-19.
110. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45(D1):D362-d8.
111. Tóth G, Pál D, Sugár S, Kovalszky I, Dezső K, Schlosser G, et al. Expression of glycosaminoglycans in cirrhotic liver and hepatocellular carcinoma-a pilot study including etiology. Anal Bioanal Chem. 2022;414(13):3837-46.
112. Tóth G, Turiák L. HPLC-MS Characterization of Tissue-Derived Heparan Sulfate and Chondroitin Sulfate. Methods Mol Biol. 2023;2619:71-90.
113. Ludwig N, Hong CS, Ludwig S, Azambuja JH, Sharma P, Theodoraki MN, et al. Isolation and Analysis of Tumor-Derived Exosomes. Curr Protoc Immunol. 2019;127(1):e91.

114. Théry C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol. 2006;Chapter 3:Unit 3.22.
115. Selman MH, Hemayatkar M, Deelder AM, Wuhrer M. Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. Anal Chem. 2011;83(7):2492-9.
116. Lageveen-Kammeijer GSM, de Haan N, Mohaupt P, Wagt S, Filius M, Nouta J, et al. Highly sensitive CE-ESI-MS analysis of N-glycans from complex biological samples. Nat Commun. 2019;10(1):2137.
117. Zhang T, Madunić K, Holst S, Zhang J, Jin C, Ten Dijke P, et al. Development of a 96-well plate sample preparation method for integrated N- and O-glycomics using porous graphitized carbon liquid chromatography-mass spectrometry. Mol Omics. 2020;16(4):355-63.
118. Kammeijer GS, Kohler I, Jansen BC, Hensbergen PJ, Mayboroda OA, Falck D, et al. Dopant Enriched Nitrogen Gas Combined with Sheathless Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry for Improved Sensitivity and Repeatability in Glycopeptide Analysis. Anal Chem. 2016;88(11):5849-56.
119. Loponte HF, Zheng J, Ding Y, Oliveira IA, Basse K, Todeschini AR, et al. GlycoGenius: a streamlined high-throughput glycan composition identification tool. Nat Commun. 2025;16(1):10335.
120. Takakura D, Harazono A, Hashii N, Kawasaki N. Selective glycopeptide profiling by acetone enrichment and LC/MS. J Proteomics. 2014;101:17-30.
121. Sugár S, Tóth G, Bugyi F, Vékey K, Karászi K, Drahos L, et al. Alterations in protein expression and site-specific N-glycosylation of prostate cancer tissues. Sci Rep. 2021;11(1):15886.
122. Skowronek P, Thielert M, Voytik E, Tanzer MC, Hansen FM, Willems S, et al. Rapid and In-Depth Coverage of the (Phospho-)Proteome With Deep Libraries and Optimal Window Design for dia-PASEF. Mol Cell Proteomics. 2022;21(9):100279.
123. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods. 2020;17(1):41-4.

124. Maxwell E, Tan Y, Tan Y, Hu H, Benson G, Aizikov K, et al. GlycReSoft: a software package for automated recognition of glycans from LC/MS data. PLoS One. 2012;7(9):e45474.
125. Downs M, Curran J, Zaia J, Sethi MK. Analysis of complex proteoglycans using serial proteolysis and EThcD provides deep N- and O-glycoproteomic coverage. Anal Bioanal Chem. 2023;415(28):6995-7009.
126. Sekiguchi R, Yamada KM. Basement Membranes in Development and Disease. Curr Top Dev Biol. 2018;130:143-91.
127. Töpfer U. Basement membrane dynamics and mechanics in tissue morphogenesis. Biol Open. 2023;12(8).
128. Li W, Deng X, Chen J. RNA-binding proteins in regulating mRNA stability and translation: roles and mechanisms in cancer. Semin Cancer Biol. 2022;86(Pt 2):664-77.
129. Qin H, Ni H, Liu Y, Yuan Y, Xi T, Li X, et al. RNA-binding proteins in tumor progression. J Hematol Oncol. 2020;13(1):90.
130. Bitaraf A, Razmara E, Bakhshinejad B, Yousefi H, Vatanmakanian M, Garshasbi M, et al. The oncogenic and tumor suppressive roles of RNA-binding proteins in human cancers. J Cell Physiol. 2021;236(9):6200-24.
131. Hu M, Ling Z, Ren X. Extracellular matrix dynamics: tracking in biological systems and their implications. J Biol Eng. 2022;16(1):13.
132. Yuan Z, Li Y, Zhang S, Wang X, Dou H, Yu X, et al. Extracellular matrix remodeling in tumor progression and immune escape: from mechanisms to treatments. Mol Cancer. 2023;22(1):48.
133. Desai N, Sahel D, Kubal B, Postwala H, Shah Y, Chavda VP, et al. Role of the Extracellular Matrix in Cancer: Insights into Tumor Progression and Therapy. Advanced Therapeutics. 2025;8(2):2400370.
134. Sugár S, Bugyi F, Tóth G, Pápay J, Kovalszky I, Tornóczky T, et al. Proteomic Analysis of Lung Cancer Types-A Pilot Study. Cancers (Basel). 2022;14(11).
135. Chen X, Shi C, He M, Xiong S, Xia X. Endoplasmic reticulum stress: molecular mechanism and therapeutic targets. Signal Transduct Target Ther. 2023;8(1):352.
136. Read A, Schröder M. The Unfolded Protein Response: An Overview. Biology (Basel). 2021;10(5).

137. Hetz C, Zhang K, Kaufman RJ. Mechanisms, regulation and functions of the unfolded protein response. Nat Rev Mol Cell Biol. 2020;21(8):421-38.
138. Ma Y, Wang L, Jia R. The role of mitochondrial dynamics in human cancers. Am J Cancer Res. 2020;10(5):1278-93.
139. Zong Y, Li H, Liao P, Chen L, Pan Y, Zheng Y, et al. Mitochondrial dysfunction: mechanisms and advances in therapy. Signal Transduct Target Ther. 2024;9(1):124.
140. Ricciardelli C, Sakko AJ, Stahl J, Tilley WD, Marshall VR, Horsfall DJ. Prostatic chondroitin sulfate is increased in patients with metastatic disease but does not predict survival outcome. Prostate. 2009;69(7):761-9.
141. Tóth G, Sugár S, Pál D, Fügedi KD, Drahos L, Schlosser G, et al. Glycosaminoglycan Analysis of FFPE Tissues from Prostate Cancer and Benign Prostate Hyperplasia Patients Reveals Altered Regulatory Functions and Independent Markers for Survival. Cancers (Basel). 2022;14(19).
142. Weyers A, Yang B, Yoon DS, Park JH, Zhang F, Lee KB, et al. A structural analysis of glycosaminoglycans from lethal and nonlethal breast cancer tissues: toward a novel class of theragnostics for personalized medicine in oncology? Omics. 2012;16(3):79-89.
143. Biskup K, Stellmach C, Braicu EI, Sehouli J, Blanchard V. Chondroitin Sulfate Disaccharides, a Serum Marker for Primary Serous Epithelial Ovarian Cancer. Diagnostics (Basel). 2021;11(7).
144. Pudełko A, Wisowski G, Olczyk K, Koźma EM. The dual role of the glycosaminoglycan chondroitin-6-sulfate in the development, progression and metastasis of cancer. Febs j. 2019;286(10):1815-37.
145. Marques C, Reis CA, Vivès RR, Magalhães A. Heparan Sulfate Biosynthesis and Sulfation Profiles as Modulators of Cancer Signalling and Progression. Front Oncol. 2021;11:778752.
146. Spyrou A, Roy A, Xiong A, Kundu S, Lu X, Jansson Y, et al. Heparan sulfate N-deacetylase/N-sulfotransferase-1 regulates glioblastoma cell migration and invasion. Matrix Biol. 2025;141:1-15.
147. Tzeng ST, Tsai MH, Chen CL, Lee JX, Jao TM, Yu SL, et al. NDST4 is a novel candidate tumor suppressor gene at chromosome 4q26 and its genetic loss predicts adverse prognosis in colorectal cancer. PLoS One. 2013;8(6):e67040.

148. Bebelman MP, Janssen E, Pegtel DM, Crudden C. The forces driving cancer extracellular vesicle secretion. Neoplasia. 2021;23(1):149-57.
149. Lopez K, Lai SWT, Lopez Gonzalez EJ, Dávila RG, Shuck SC. Extracellular vesicles: A dive into their role in the tumor microenvironment and cancer progression. Front Cell Dev Biol. 2023;11:1154576.
150. Zhang H, Hui D, Fu X. Roles of Fibulin-2 in Carcinogenesis. Med Sci Monit. 2020;26:e918099.
151. Ma Y, Nenkov M, Schröder DC, Abubrig M, Gassler N, Chen Y. Fibulin 2 Is Hypermethylated and Suppresses Tumor Cell Proliferation through Inhibition of Cell Adhesion and Extracellular Matrix Genes in Non-Small Cell Lung Cancer. Int J Mol Sci. 2021;22(21).
152. Moon PG, Lee JE, Cho YE, Lee SJ, Chae YS, Jung JH, et al. Fibronectin on circulating extracellular vesicles as a liquid biopsy to detect breast cancer. Oncotarget. 2016;7(26):40189-99.
153. Ye X, Ling B, Xu H, Li G, Zhao X, Xu J, et al. Clinical significance of high expression of proliferating cell nuclear antigen in non-small cell lung cancer. Medicine (Baltimore). 2020;99(16):e19755.
154. Sun DS, Chang HH. Extracellular vesicles: Function, resilience, biomarker, bioengineering, and clinical implications. Tzu Chi Med J. 2024;36(3):251-9.
155. Couchman JR, Pataki CA. An introduction to proteoglycans and their localization. J Histochem Cytochem. 2012;60(12):885-97.
156. Wang T, Liu X, Tian Q, Liang T, Chang P. Reduced SPOCK1 expression inhibits non-small cell lung cancer cell proliferation and migration through Wnt/β-catenin signaling. Eur Rev Med Pharmacol Sci. 2018;22(3):637-44.
157. Petővári G, Tóth G, Turiák L, A LK, Pálóczi K, Sebestyén A, et al. Dynamic Interplay in Tumor Ecosystems: Communication between Hepatoma Cells and Fibroblasts. Int J Mol Sci. 2023;24(18).
158. Onyeisi JOS, Lopes CC, Götte M. Syndecan-4 as a Pathogenesis Factor and Therapeutic Target in Cancer. Biomolecules. 2021;11(4).
159. Manon-Jensen T, Itoh Y, Couchman JR. Proteoglycans in health and disease: the multiple roles of syndecan shedding. Febs j. 2010;277(19):3876-89.
160. Hanahan D. Hallmarks of cancer-Then and now, and beyond. Cell. 2026.

161. Sobanski T, Rose M, Suraweera A, O'Byrne K, Richard DJ, Bolderson E. Cell Metabolism and DNA Repair Pathways: Implications for Cancer Therapy. Front Cell Dev Biol. 2021;9:633305.
162. Li Y, Wang J, Chen W, Lu H, Zhang Y. Comprehensive review of MS-based studies on N-glycoproteome and N-glycome of extracellular vesicles. Proteomics. 2024;24(11):e2300065.
163. Thomas D, Rathinavel AK, Radhakrishnan P. Altered glycosylation in cancer: A promising target for biomarkers and therapeutics. Biochim Biophys Acta Rev Cancer. 2021;1875(1):188464.
164. Giurini EF, Pappas SG, Gupta KH. Sweet Surprises: Decoding Tumor-Associated Glycosylation in Cancer Progression and Therapeutic Potential. Cells. 2026;15(3).
165. Wight TN, Kinsella MG, Evanko SP, Potter-Perigo S, Merrilees MJ. Versican and the regulation of cell phenotype in disease. Biochim Biophys Acta. 2014;1840(8):2441-51.
166. Hohenester E, Yurchenco PD. Laminins in basement membrane assembly. Cell Adh Migr. 2013;7(1):56-63.
167. Koh Kok JY, Goh BH, Lee WL. Altered glycosylation of EVs in cancers: galectin-3-binding protein (LGALS3BP). Discov Oncol. 2025;16(1):1336.
168. Papadas A, Arauz G, Cicala A, Wiesner J, Asimakopoulos F. Versican and Versican-matrikines in Cancer Progression, Inflammation, and Immunity. J Histochem Cytochem. 2020;68(12):871-85.
169. Qin Y, Rodin S, Simonson OE, Hollande F. Laminins and cancer stem cells: Partners in crime? Semin Cancer Biol. 2017;45:3-12.
170. Nonnast E, Mira E, Mañes S. The role of laminins in cancer pathobiology: a comprehensive review. J Transl Med. 2025;23(1):83.
171. Capone E, Iacobelli S, Sala G. Role of galectin 3 binding protein in cancer progression: a potential novel therapeutic target. J Transl Med. 2021;19(1):405.
172. Bastian K, Scott E, Elliott DJ, Munkley J. FUT8 Alpha-(1,6)-Fucosyltransferase in Cancer. Int J Mol Sci. 2021;22(1).
173. van Vliet SJ, van Kooyk Y. Sialic acids in cancer biology and immunity-recent advancements. J Biol Chem. 2025;301(10):110641.

174. Pendiuk Goncalves J, Walker SA, Aguilar Díaz de León JS, Yang Y, Davidovich I, Busatto S, et al. Glycan Node Analysis Detects Varying Glycosaminoglycan Levels in Melanoma-Derived Extracellular Vesicles. Int J Mol Sci. 2023;24(10).
175. Walker SA, Aguilar Díaz De León JS, Busatto S, Wurtz GA, Zubair AC, Borges CR, et al. Glycan Node Analysis of Plasma-Derived Extracellular Vesicles. Cells. 2020;9(9).
176. Xiong DD, Li JD, He RQ, Li MX, Pan YQ, He XL, et al. Highly expressed carbohydrate sulfotransferase 11 correlates with unfavorable prognosis and immune evasion of hepatocellular carcinoma. Cancer Med. 2023;12(4):4938-50.
177. Zhang P, Chen D, Cui H, Luo Q. High Expression of CHST11 Correlates with Poor Prognosis and Tumor Immune Infiltration of Pancreatic Cancer. Clin Lab. 2022;68(12).
178. Khazamipour N, Al-Nakouzi N, Oo HZ, Ørum-Madsen M, Steino A, Sorensen PH, et al. Oncofetal Chondroitin Sulfate: A Putative Therapeutic Target in Adult and Pediatric Solid Tumors. Cells. 2020;9(4).
179. Spliid CB, Toledo AG, Sanderson P, Mao Y, Gatto F, Gustavsson T, et al. The specificity of the malarial VAR2CSA protein for chondroitin sulfate depends on 4-O-sulfation and ligand accessibility. J Biol Chem. 2021;297(6):101391.
180. Oo HZ, Lohinai Z, Khazamipour N, Lo J, Kumar G, Pihl J, et al. Oncofetal Chondroitin Sulfate Is a Highly Expressed Therapeutic Target in Non-Small Cell Lung Cancer. Cancers (Basel). 2021;13(17).
181. Salanti A, Clausen TM, Agerbæk M, Al Nakouzi N, Dahlbäck M, Oo HZ, et al. Targeting Human Cancer by a Glycosaminoglycan Binding Malaria Protein. Cancer Cell. 2015;28(4):500-14.
182. Enciso-Martinez A, Løppke C, Koene JJ, Ebbelaar J, van der Geest M, Los M, et al. Targeting Oncofoetal Chondroitin Sulphate Allows Identification of Tumour-Derived Extracellular Vesicles. J Extracell Vesicles. 2025;14(6):e70106.
183. Zhao Y, Wen C, Wang Q, Qing Y, Tondi S, Reina C, et al. Use of the Malaria Protein VAR2CSA for the Detection of Small Extracellular Vesicles to Diagnose Adenocarcinoma. J Extracell Vesicles. 2025;14(4):e70067.

9. Bibliography of the candidate's publications

9.1. Original publications related to the dissertation

1. M. Balbisi, T. Langó, V.N. Horváth, D. Pál, G. Schlosser, G. Kecskeméti, Z. Szabó, K. Ilyés, N. Nagy, O. Tóth, J. Zheng, G.S.M. Lageveen-Kammeijer, T. Visnovitz, Z. Varga, B.G. Vértessy, L. Turiák: Distinct proteomic and glycosylation signatures differentiate A549 tumor and BEAS-2B nontumor cell line–derived small extracellular vesicles. *Molecular & Cellular Proteomics* 2026, 25(3): 101524.

   IF: 5.5; D1

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

   IF: 3.8; D1

9.2. Review publication related to the dissertation

3. M. Balbisi, S. Sugár, L. Turiák: Protein glycosylation in lung cancer from a mass spectrometry perspective. *Mass Spectrometry Reviews* 2026, 45(3): 455-475.

   IF: 6.6; D1

9.3. Publications unrelated to the dissertation

4. D. Lenzinger, L. Lankovics, I. Dudás, T. Bárkai, Zs. Szász, M. Balbisi, A. Y. Wu, K. V. Vukman, K. Fletcher, A. Csomos, Z. Mucsi, E. Bugyik, Cs. Cserép, Á. Dénes, Sz. Bősze, L. Turiák, C. P. Lai, E. I. Buzás, T. Visnovitz: Stress-induced switch in small extracellular vesicle secretion: from constitutive "torn bag mechanism" to exocytosis. *Journal of Extracellular Vesicles* 2026, in press.

   IF: 21.7; D1

5. D. Virág, K. M. Karvaly, A. Molnár, M. Balbisi, J. Németh, L. Turiák: Mass spectrometry insights into post-translational modifications in extracellular vesicles. *Mass Spectrometry Reviews* 2026, in press.

   IF: 6.6; D1

6. F. Bugyi, M. Balbisi, S. Sugár, L. Váncza, E. Regős, I. Kovalszky, I. Laczó, T. Harkó, G. Kecskeméti, Z. Szabó, J. Moldvay, L. Drahos, L. Turiák: Unveiling unique protein and phosphorylation signatures in lung adenocarcinomas with and without ALK, EGFR, and KRAS genetic alterations. *Molecular Oncology* 2025, 19: 3243-3265.

IF: 4.5; D1

7. B. György, R. Szatmári, T. Ditrói, F. Torma, K. Pálóczi, M. Balbisi, T. Visnovitz, E. Koltai, P. Nagy, E.I. Buzás, S. Horvath, Zs. Radak: The protein cargo of extracellular vesicles correlates with the epigenetic aging clock of exercise sensitive DNAmFitAge. *Biogerontology* 2025, 26: 35.

   IF: 4.1; Q1

8. B. György, K. Pálóczi, M. Balbisi, L. Turiák, L. Drahos, T. Visnovitz, E. Koltai, Zs. Radák: Effect of the 35 nm and 70 nm size exclusion chromatography (SEC) column and plasma storage time on separated extracellular vesicles. *Current Issues in Molecular Biology* 2024, 46(5): 4337-4357.

   IF: 3.0; Q2

9. B. Szeitz, T. Glasz, Z. Herold, G. Tóth, M. Balbisi, J. Fillinger, Sz. Horváth, R. Mohácsi, H.J. Kwon, J. Moldvay, L. Turiák, A.M. Szász: Spatially resolved proteomic and transcriptomic profiling of anaplastic lymphoma kinase-rearranged pulmonary adenocarcinomas reveals key players in inter- and intratumoral heterogeneity. *International Journal of Molecular Sciences* 2023, 24(14): 11369.

   IF: 4.9; D1

10. G. Tóth, S. Sugár, M. Balbisi, B.A. Molnár, F. Bugyi, K.D. Fügedi, L. Drahos, L. Turiák: Optimized sample preparation and microscale separation methods for high-sensitivity analysis of hydrophilic peptides. *Molecules* 2022, 27(19): 6645.

    IF: 4.6; Q1

11. M. Balbisi, R.A. Horváth, M. Szőri, P. Jedlovszky: Computer simulation investigation of the adsorption of acetamide on low density amorphous ice. An astrochemical perspective. *The Journal of Chemical Physics* 2022, 156: 184703.

    IF: 4.4; Q1

12. M. Balbisi, R.A. Horváth, M. Szőri, P. Jedlovszky: Adsorption of acetamide on crystalline and amorphous ice under atmospheric conditions. A grand canonical Monte Carlo simulation study. *Journal of Molecular Liquids* 2022, 354: 118870.

    IF: 6.0; Q1

13. S. Góbi, M. Balbisi, Gy. Tarczay: Local and remote conformational switching in 2-fluoro-4-hydroxy benzoic acid. *Photochem* 2022, 2(1): 102-121.

    IF: -

10. Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Lilla Turiák, for her guidance, support, and dedication throughout my work. I am especially grateful for the knowledge and experience I have gained under her supervision, which have greatly contributed to my development and have significantly shaped my approach to research.

I am grateful to my former co-supervisor, Gábor Tóth, whose guidance during my MSc studies provided valuable professional and personal experience.

I would like to thank László Drahos, whose lectures on mass spectrometry first drew my attention to the research carried out at HUN-REN TTK, leading me to join the group.

Over the years, I had the opportunity to supervise BSc and MSc students, which has been an incredibly enriching experience. I am thankful to my students – Boglárka Cziráki, Virág Nikolett Horváth, Viktória Kiss, and Alexandra Molnár – for their dedication and hard work.

I am grateful to all current and former members of the MTA-HUN-REN TTK Lendület (Momentum) Glycan Biomarker Research Group and the MS Proteomics Research Group for the supportive and motivating research environment and the valuable scientific discussions: András Ács, Fanni Bugyi, Rachma Dessidianti, Kata Fügedi, Márkó Grabarics, Ágnes Gömöry, Krisztián Karvaly, Salvatore Laudani, Balázs Molnár, Júlia Németh, Kinga Nagy, Domonkos Pál, Martin Papp, Ágnes Révész, Dániel Szabó, Simon Sugár, Károly Vékey, and Dávid Virág.

I had the opportunity to complete two internships that contributed significantly to both my professional and personal development. I would like to express my gratitude to Guinevere Lageveen-Kammeijer and the Analytical Biochemistry Research Group at the University of Groningen, as well as to Martina Marchetti-Deschmann, Samuele Zoratto, and the Mass Spectrometric Bio- and Polymer Analysis Research Group at the Technical University of Vienna.

I would like to extend my special thanks to those who contributed to the work presented in this dissertation: Bella Bruszel, János Fillinger, Kinga Ilyés, Gábor Kecskeméti, Tamás Langó, Judit Moldvay, Nikolett Nagy, Armel Nicholas, Gitta Schlosser, Zoltán Szabó, A. Marcell Szász, Beáta Szeitz, Otília Tóth, Zoltán Varga, Tamás Visnovitz, Beáta

Vértessy, and Jing Zheng. I am also grateful to all colleagues with whom I have worked on other projects.

Before committing to proteomics and glycomics, I explored two other research areas. I am thankful to my former supervisors and colleagues: György Tarczay, István Pál Csonka, and the ELTE Laboratory of Molecular Spectroscopy, as well as Pál Jedlovszky and Réka Anna Horváth. The knowledge and skills I acquired during this period were of great help to my research.

I am grateful to the Doctoral School of Semmelweis University, Pharmaceutical Sciences and Health Technologies Division, for providing me the opportunity to pursue my PhD, and to the Institute of Chemistry at Eötvös Loránd University, where I completed my BSc and MSc degrees, for laying the foundation of my scientific education.

I would also like to thank the teachers at Jedlik Ányos Secondary School for the profound impact they have had on me. I am especially grateful to my chemistry teacher, Beatrix Elekné Becz, whose early encouragement and guidance fostered my interest in chemistry and shaped my scientific path.

Last but not least, I would like to express my sincere gratitude to my family and friends for their continuous support and encouragement, without which this work would not have been possible.