DOI:10.14753/SE.2023.2877

SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

2877.

SUGÁR SIMON NÁNDOR

A gyógyszerészeti tudományok korszerű kutatási irányai
című program

Programvezető: Dr. Antal István, egyetemi tanár

Témavezető: Dr. Turiák Lilla, tudományos munkatárs

DOI:10.14753/SE.2023.2877

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

PhD thesis

Simon Nándor Sugár

Doctoral School of Pharmaceutical Sciences
Semmelweis University

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

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

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

Budapest
2023

DOI:10.14753/SE.2023.2877

Table of Contents

List of Abbreviations 4

1. Introduction 7

1.1. Cancer 7

1.1.1. Prostate Cancer 7

1.1.2. Lung Cancer 8

1.2. Proteomics 9

1.2.1. Proteins and *N*-glycosylation 9

1.2.2. Approaches in Proteomics 11

1.2.3. Approaches in *N*-glycoproteomics 11

1.3. Mass Spectrometry-based Proteomics 12

1.3.1. Ion Sources Used in Proteomics 13

1.3.2. Mass Analyzers Used in Proteomics 14

1.3.3. Tandem Mass Spectrometry and Fragmentation Methods 15

1.3.3.1. Fragmentation of Peptides and *N*-glycopeptides 17

1.3.4. Liquid Chromatography Coupled Mass Spectrometry 18

1.3.5. Mass Spectrometry Data Acquisition 19

1.3.6. Protein and Glycoprotein Identification 20

1.3.7. Quantitation in Proteomics 20

1.4. Proteomics in Cancer Research 21

1.4.1. Prostate Cancer Proteomics 21

1.4.2. Lung Cancer Proteomics 22

1.5. Special Considerations in the Analysis of FFPE Tissue Samples 23

2. Objectives 24

3. Methods 25

3.1. Materials 25

3.2. Patient and Sample Information – Prostate Tissue 25
3.3. Patient and Sample Information – Lung Tissue 25
3.4. Sample Preparation 26
3.4.1. On-surface Digestion 26
3.4.2. Reversed-phase Purification 27
3.4.3. Glycopeptide Enrichment for Prostate TMAs 27
3.5. nanoUHPLC-MS(MS) Analysis 27
3.6. Software Used 28
3.7. Data Analysis 28
3.7.1. Prostate Cancer Proteomics and Glycoproteomics 28
3.7.2. Lung Cancer Proteomics 29
4. Results 30
4.1. Prostate Cancer Proteomics and *N*-glycosylation 30
4.1.1. Differences between Normal and Cancerous Tissue 31
4.1.2. Differences with Prostate Cancer Grade Progression 33
4.1.3. Biological Processes Altered in Prostate Cancer 34
4.2. Lung Cancer Proteomics 36
4.2.1. General Differences between Tumor and Adjacent Normal Tissue 37
4.2.2. Lung Cancer Type-specific Differences between Tumor and Adjacent Normal Tissue 38
4.2.3. Differences between the Different Types of Tumor Tissue 40
5. Discussion 43
5.1. Proteomics and *N*-glycoproteomics of Different Grades of Prostate Cancer 43
5.2. Proteomics of the Four Lung Cancer Types 46
6. Conclusions 54
6.1. Proteomics and *N*-glycoproteomics of Different Grades of Prostate Cancer 54

DOI:10.14753/SE.2023.2877

6.2. Proteomics of the Four Lung Cancer Types 54

7. Summary 55

8. References 56

9. Bibliography of the candidate's publications 76

9.1. Publications related to the dissertation 76

9.2. Publications unrelated to the dissertation 76

10. Acknowledgements 79

DOI:10.14753/SE.2023.2877

List of Abbreviations

ABC	ammonium bicarbonate
AC	adenocarcinoma
ACLY	ATP-citrate synthase
ACN	acetonitrile
ACT	actins
ALK	anaplastic lymphoma kinase
ANK1	ankyrin-1
ANX	annexins
APO	apolipoproteins
AR	antigen retrieval
AUC	area under curve
CAPRA	Cancer of the Prostate Risk Assessment
CID	collision-induced dissociation
CO6A2	collagen alpha-2(VI) chain
CRPC	castration-resistant prostate cancer
CTLA-4	cytotoxic T-lymphocyte-associated antigen 4
DDA	data dependent acquisition
DIA	data independent acquisition
DRE	digital rectal examination
ECM	extracellular matrix
EGFR	epidermal growth factor receptor
EIF	eukaryotic translation initiation factors
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
ESI	electrospray ionization
ETD	electron transfer dissociation
FA	formic acid
FC	fold change
FDR	false discovery rate
FF	fresh frozen
FFPE	formalin fixated paraffin embedded
FIB	fibrinogens
FSCN1	fascin
FT-ICR	Fourier-transform ion cyclotron resonance
GO	Gene Ontology
GOBP	gene ontology biological process
GOCC	gene ontology cellular compartment
GOMF	gene ontology molecular function
GSEA	Gene Set Enrichment Analysis

HCD	higher-energy collision dissociation
HFBA	heptafluorobutyric acid
HILIC	hydrophilic interaction liquid chromatography
HPLC	high performance liquid chromatography
IF	eukaryotic translation initiation factors
IGG1	immunoglobulin gamma-1 heavy chain
IGHG2	immunoglobulin heavy constant gamma 2
IQR	inter-quartile range
KEGG	Kyoto Encyclopedia of Genes and Genomes
KRAS	Kirsten rat sarcoma viral oncogene homolog
LAM	laminins
LC	lung cancer
LCC	large cell carcinoma
LCNEC	large cell neuroendocrine carcinoma
LEG	galectins
LFQ	label-free quantitation
MALDI	matrix-assisted laser desorption ionization
MAP	microtubule-associated proteins
MFAP4	microfibril-associated glycoprotein 4
MMP	matrix metalloproteinases
mRNA	messenger RNA
MS	mass spectrometry
MTORC1	mammalian target of Rapamycin 1
nanoUHPLC	nanoflow HPLC
NES	Normalized Enrichment Score
NID	nidogens
NSCLC	non-small cell lung cancer
PARP	poly(ADP)-ribose polymerase
PCa	prostate cancer
PCA	principal component analysis
PD-1	programmed death-ligand 1 receptor
PGC	porous graphitic carbon
PGS1	biglycan
PNGaseF	peptide-*N*-glycosidase F
POSTN	periostin
PPAP	prostatic acid phosphatase
PPI	protein-protein interaction
PSA	prostate-specific antigen
PSM	peptide-spectra match
PTM	post-translational modification

Q	quadrupole
RP	reversed-phase
RT	retention time
SCLC	small cell lung cancer
SPE	solid phase extraction
SPT	spectrins
SqCC	squamous cell carcinoma
TBB	tubulin beta chains
TFA	trifluoroacetic acid
THRB	prothrombin
TMA	tissue microarray
TME	tumor microenvironment
TOF	time-of-flight
tRNA	transfer RNA

1. Introduction

1.1. Cancer

Cancer is one of the leading causes of death worldwide. According to the World Health Organization, in 2019, it was the first or second leading cause of death in 112 of 183 countries for people under the age of 70. [1] In 2020, there were an estimated 19.3 million new cases and 10 million cancer-related deaths worldwide. [2] There is a multitude of different cancers; they are usually named after the body part of origin (e.g., prostate cancer).

The diagnosis of cancer is different for the distinct cancer types; however, it always involves **histological grading** and **pathological staging**. Grading is based on the appearance of cancer cells under the microscope, according to their level of differentiation: Grade 1 (low grade, well differentiated), Grade 2 (intermediate grade, moderate differentiation), Grade 3 (high grade, poor differentiation). Staging is based on several factors that describe the severity of the disease. The most commonly used system, the TNM system reports on: T—the size of the primary tumor, N—whether it has spread to nearby lymph nodes, and M—whether it has metastasized. [3–5]

1.1.1. Prostate Cancer

Prostate cancer (PCa) has the second highest incidence (14.1% of cases) and the fifth highest mortality (6.8% of cancer-related deaths) out of the different cancer types in males worldwide. [2] As these numbers indicate, PCa has a relatively low mortality compared to other types of cancer: the five-year survival of patients with early-stage PCa is almost 100%. On the other hand, for advanced cases—which usually metastasized to the bones, liver, or lungs—it is 30%. [6] The risk assessment of PCa considers multiple factors, e.g., grading, staging, age, etc.

Grading is performed by a pathologist, and can be based on different kinds of features: **histological grading** is based on levels of cell differentiation, ranging from 1–3 (see *1.1 Cancer*); **Gleason grading** considers not only cell appearances but tissue architecture as well, and considers both the least and second least differentiated patterns, ranges from 1–5; ISUP (International Society of Urological Pathology) grading is a slightly modified, updated Gleason system, also ranging from 1–5; while the CAPRA (Cancer of the Prostate Risk Assessment), and D'Amico (developed by Anthony V. D'Amico, MD,

PhD) systems combine Gleason grades with other clinical parameters, designed for more accurate PCa risk prediction. [7–10]

PCa screening is based on measuring the levels of prostate-specific antigen (PSA) in blood, and digital rectal examination (DRE). [11–13] Further diagnosis is usually made using targeted biopsies. [14,15] Therapy options in PCa include active surveillance or observation for patients with low-risk PCa; surgery or radiation therapy for localized PCa; hormone therapy; chemotherapy; immunotherapy or immune checkpoint inhibitors; targeted therapy for patients with known mutations (e.g., poly(ADP)-ribose polymerase (PARP) inhibitors); and therapies specific to bone metastasis (e.g., bisphosphonates, corticosteroids, external radiation therapy, or radiopharmaceuticals). [16]

1.1.2. Lung Cancer

Lung cancer (LC) has the second highest incidence (11.4% of new diagnoses) and the highest mortality (18.0% of deaths) out of all cancer types worldwide. [2] The large numeric difference between these two metrics is due to the very low five-year survival rate of the disease (10-20% in most countries), [2] which is mostly because a large part of cases is already metastatic at the time of diagnosis. [17,18] In publications focused on global LC incidence and mortality, Hungary is usually reported with the highest numbers; however, this is reportedly an artifact due to differences in data collection methodologies. In reality, they are not higher than the average of other Central European countries. [19]

Lung cancer is a highly heterogeneous disease. It is classified into two main types: **small cell lung cancer** (SCLC, 15% of cases) and **non-small cell lung cancer** (NSCLC, 85% of cases) based on their appearance under the microscope. NSCLC can further be classified into three major subtypes based on histology: **adenocarcinoma** (AC, 45% of cases), **squamous cell carcinoma** (SqCC, 25% of cases), and **large cell carcinoma** (LCC, 15% of cases). [20]

The classification of LC is a very important part of establishing a diagnosis, as depending on the type or subtype, the available treatment options and prognosis might be different. SCLC is generally considered more aggressive than NSCLC due to the rapid growth of cancer cells and its high tendency to metastasize to other parts of the body, which also rules out the possibility of surgery in most cases. [21] Treatment options for SCLC generally only include chemotherapy and radiation therapy, while for NSCLC, surgery, and—depending on the subtype and certain genetic and protein markers—targeted

therapies and immunotherapy are also available. [22] These key differences between the two types of LC are reflected in their respective five-year survival rates, which is much higher for patients with NSCLC in all metastatic stages of the disease.

1.2. Proteomics

The definition of **proteomics** has changed significantly over the last decades, as many new technologies and methodologies emerged expanding possibilities. It is the study of the **proteome**—the proteins expressed by a cell or organism, their quantity, structure, function, post-translational modifications, and interactions. [23–25]

1.2.1. Proteins and *N*-Glycosylation

Proteins are large macromolecules that consist of amino acid residues (21 different proteinogenic amino acids for humans). Their biosynthesis is template-driven and can be separated into two phases: transcription and translation. During transcription the information encoded in a gene is transferred to messenger RNA (mRNA) with the help of RNA polymerases in the nucleus of the cell. The mRNA is then transported to the cytosol, where translation is carried out by ribosomes based on the information provided by the mRNA in the presence of transfer RNA (tRNA) that carry the necessary amino acid building blocks. After translation, many proteins undergo chemical modifications called post-translational modification (PTM). **Glycosylation**—the addition of polysaccharide unit(s) (glycans) to the protein—is considered the most common PTM, which happens in the endoplasmic reticulum (ER) and the Golgi apparatus. The two most common types are *O*- and ***N*-linked glycosylation**. [26]

In the case of *N*-glycosylation the glycan can be attached to the amide nitrogen of the asparagine residue in a Ser/Thr-X-Asn sequence (where X is any amino acid except for proline) called a **glycosylation site** (also referred to as glycosite), which is usually referred to by an “N” (for asparagine) followed by the number in sequence (e.g., N294). The process starts in the ER, where a dolichol-phosphate-linked precursor is transferred to the Asn with the help of an oligosaccharyltransferase. The next step is the processing and maturation of the glycan in the Golgi. Here, large part of the polysaccharide is removed with the help of glycosidases, then new sugar residues are added by glycosyltransferases resulting in three types of mature glycan structures: **high-mannose** (or oligomannose), **hybrid**, and **complex**. All of them have a common, conserved core structure and can only contain a select number of different monosaccharides (*Figure 1*).

[26] Other, less common glycans include bisecting, paucimannose, and LacdiNAc structures.

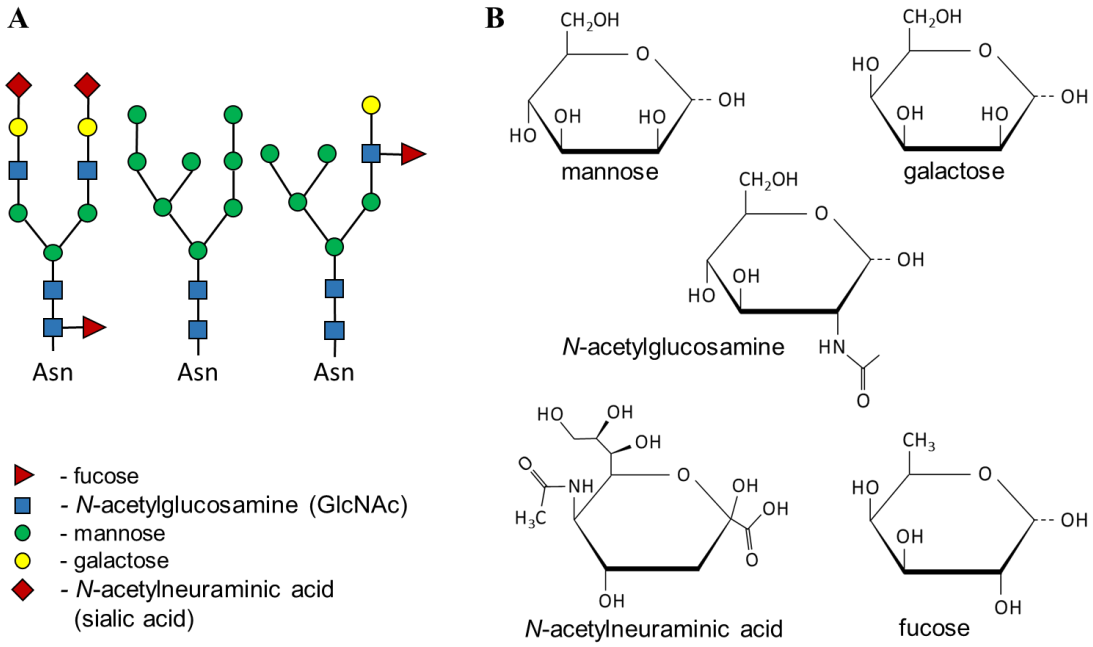

Figure 1: *N*-glycan types common in humans. (**A**) The three glycan types from left to right: complex, high-mannose, and hybrid. (**B**) Monosaccharides common in human *N*-glycans. Figure made by the author using Microsoft PowerPoint.

A glycoprotein can have multiple glycosylation sites, and usually present several different glycan structures simultaneously—they have several different **glycoforms**. Glycosylation can be characterized by two types of heterogeneity, **macroheterogeneity** refers to glycosite occupancy, while **microheterogeneity** refers to the collection of different glycoforms. Glycosylation can alter the structure and subsequently the function of a protein by a large degree, therefore has various important biological functions, most notably, glycoproteins are heavily involved in cell-cell communication. Glycans presented on the cell surface affect signal transduction, molecular recognition, [27] and the entry of viruses into the cell—for example the SARS-CoV-2 virus uses its heavily glycosylated spike proteins to bind to angiotensin-converting enzyme 2 (an *N*-glycoprotein) cell surface receptors to facilitate cell entry. [28] Aberrant glycosylation has been repeatedly shown to be of diagnostic value for multiple diseases, while

monoclonal antibodies used for the treatment of several diseases, e.g., cancer are also glycosylated. [29,30]

1.2.2. Approaches in Proteomics

In proteomics, there are currently two major methodologies used for the analysis of proteins. In the **bottom-up** methodology (often called shotgun proteomics), samples are first subjected to proteolysis by proteases (most frequently trypsin) that cleave proteins at specific amino acid residues resulting in peptides. These are then predominantly analyzed using mass spectrometry following chromatographic separation, and proteins are identified and quantified through their respective peptides. [25,31–34] Contrary, in the **top-down** methodology proteins are analyzed in their intact form. Additionally, to the two approaches mentioned above, middle-down technique is also used in specific cases (e.g., the analysis of PTMs on monoclonal antibodies), where proteins are only partially digested. [35–37]

The two major methodologies have different applications. The main difference is that bottom-up proteomics is an older, well-established approach, that allows for the high-throughput analysis (identification and quantitation) of a very large number of proteins in parallel; while top-down proteomics is a more recent technology, that enables the investigation of different proteoforms for less complex samples and requires more expensive instrumentation and expertise, also the analysis of proteins with large molecular weights is problematic. Although there have been significant improvements in top-down methodologies in recent years—making its breadth of applications wider, [38] bottom-up proteomics remains the most often used approach, and the cornerstone of proteomics.

1.2.3. Approaches in *N*-Glycoproteomics

The analysis of *N*-glycoproteins often requires different, or modified methodologies than proteins. This is due to the previously described (*1.2.1. Proteins and N-glycosylation*) glycan structure that has very different molecular characteristics than the peptide backbone and has a highly variable structure, thus increasing sample complexity. Currently, *N*-glycoproteomics is mostly based on two well-established approaches, **released glycan analysis** and **intact glycopeptide analysis**.

In released glycan analysis, the peptides (or less frequently proteins) are subjected to an *N*-glycosidase enzyme—usually peptide-*N*-glycosidase F (PNGaseF)—that cleaves the

glycan from the peptide backbone, which allows for the separate analysis of the two types of molecular sets. After de-glycosylation, they are usually separated based on the difference in their hydrophobicity/hydrophilicity, charge, size, or affinity to specific molecules (e.g., lectins). The advantages of this method are that different methods can be used for the analysis of glycans and peptide backbones; the sample complexity is much smaller; more detailed structure elucidation is possible; sensitivity is less of an issue (usually with help of derivatization), as glycans from different glycosylation sites are detected together. The disadvantage of this method is that it only provides averaged information on the different glycosylation sites (and subsequently glycoproteins) studied, while site-specific information is lost. [39–41]

Intact glycopeptide analysis—as the name suggests, is the direct analysis of glycopeptides after enzymatic digestion. Contrary to released glycan analysis, site-specific information is retained; however, it comes with a number of disadvantages. Sample complexity is much higher, and signals can be much smaller, as every glycan-peptide combination is a separate species (with a great dynamic range) opposed to released glycan analysis. Furthermore, glycosylation decreases the ionization efficiency of peptides [42] which can also lead to the suppression of glycopeptides in the ion source (ion suppression) by unmodified peptides. [43] To address these problems the **enrichment of glycopeptides** is paramount, with several different approaches currently being used.

The two main approaches for glycopeptide enrichment are precipitation-based and chromatography-based methods. Precipitation-based methods make use of the different solubility of the two types of molecular sets in the mix of water and a polar organic solvent, e.g., in acetone or ethanol. Chromatography-based methods are usually used in a solid phase extraction (SPE) framework, where the stationary phase is packed in a cartridge or a pipette tip. Several different stationary phases are used, most commonly: hydrophilic interaction liquid chromatography (HILIC), porous graphitic carbon (PGC), or lectin-based affinity sorbents. [44–48]

1.3. Mass Spectrometry-Based Proteomics

As complex organisms can express a vast number of different proteins—for human cells an estimated 100 000, which is further diversified by different proteoforms—with great dynamic range in their abundance, great differences in molecular size and chemical properties; there is a multitude of different analytical techniques used in proteomics (e.g.,

gel-based methods, western blotting, enzyme-linked immunosorbent assays (ELISA), etc.). Currently, however, mass spectrometry-based approaches are the most widely used. Since the development of soft ionization techniques, such as **electrospray ionization** (ESI) and **matrix-assisted laser desorption ionization** (MALDI) by Fenn, [49,50] and Hillenkamp and Karas [51] respectively (a similar technique was applied to proteins by Tanaka [52]), mass spectrometry (MS) has become an indispensable tool for the analysis of proteins. The importance of these techniques is highlighted by the fact that in 2002 Fenn and Tanaka shared half of the Nobel Prize in Chemistry for their work. MS is at the center of proteomics research as it is a versatile tool that allows the high-throughput and sensitive analysis of thousands of proteins in parallel. Most of the time, it is used not as a standalone technique, but in combination with separation methods such as liquid chromatography [32] or less frequently capillary electrophoresis. [53] In further sections of the *Introduction*, when proteomics (and glycoproteomics) methods are discussed, only bottom-up applications are considered.

Mass spectrometry is an analytical technique based on the measurement of the mass to charge ratio (*m*/*z*) of ionized molecules. An MS instrument is comprised of three main parts: the ion source—where analyte ions are formed, the mass analyzer—where the ions are separated based on their *m*/*z*, and the detector. The ions are transported between these segments by electrostatic fields called ion optics. During analysis a **mass spectrum** is produced that contains the *m*/*z* of ions on the *x* axis, and their *intensity* (the signal produced in the detector) on the *y* axis.

State of the art mass spectrometers currently used in proteomics employ different technologies and have different layouts. This involves the use of various ion sources and mass analyzers, or specific additional parts.

1.3.1. Ion Sources Used in Proteomics

As previously mentioned, the development of soft ionization techniques was central for the application of MS to the analysis of proteins. Hard ionization techniques can cause the analyte molecule to fragment, which can be detrimental to the analysis of proteins or peptides, as no molecule ion is formed. Furthermore, producing gas-phase ions of large biomolecules was previously not possible, which is why John B. Fenn's Nobel Lecture was titled *"Electrospray Wings for Molecular Elephants"*. [54]

MALDI is a so called “energy-sudden” method, [54] that involves the use of a laser beam to evaporate molecules from the sample surface with the help of a matrix that has a high absorptivity at the laser’s wavelength (e.g., 2,5-dihydroxy benzoic acid). [55] During MALDI mainly singly charged ions are formed. [51,52,56,57] Advantages of the technique include its ability to ionize molecules with very high molecular weights (> 200 000 Da), its high sensitivity (up to 10^{-21} mol), its high tolerance for the presence of salts, and the ability to study mixtures directly. The main disadvantages are the difficulty of quantitation and its inability to be connected on-line to a chromatographic system. A MALDI source is usually connected to a time-of-flight-based system.

In ESI the analytes are in liquid form. The samples are solubilized in water mixed with a volatile organic solvent (e.g., acetonitrile, or methanol) in the presence of an acid or base (e.g., formic acid in positive, ammonia in negative modes), then passed through a capillary that is under high voltage. An aerosol of charged droplets is formed with the help of high temperatures and a nebulizer gas. Subsequently, through solvent evaporation Coulomb fission occurs repeatedly. [49,50,54,58,59] The three widely accepted theories explaining the final part of the ionization process—where gas-phase ions are formed from the droplets—are, for smaller molecules the ion evaporation, for larger biomolecules the charge residue model, and the chain ejection model. [60] During ESI multiply charged molecular ions can be produced, which makes the analysis of large biomolecules possible with mass analyzers operating in lower mass ranges as well. For proteomics, it is usually combined with a quadrupole time-of-flight, or an Orbitrap-based system. The main advantage of the ESI source is the possibility of coupling with liquid chromatography. [61] It is also highly sensitive (up to 10^{-18} mol) and suitable for quantitation. Disadvantages include its low tolerance towards non-volatile buffers and high salt concentrations, and its relatively high maintenance. [62]

1.3.2. Mass Analyzers Used in Proteomics

Mass analyzers are capable of separating ions based on their *m*/*z* values. The most frequently used mass analyzers in proteomics are the **quadrupole** (Q), the **time-of-flight** (TOF), the **ion trap**, the **Orbitrap**, and the **Fourier-transform ion cyclotron resonance** (FT-ICR).

The quadrupole is comprised of four cylindrical metal rods. The opposing rod-pairs are connected to each other electrically, and a radio frequency alternating current is applied

to them, with a direct current offset voltage between them. Only ions with a certain *m/z* can pass through the rods when the voltages are chosen correctly. As the quadrupole has a relatively small mass range (up to 3000 Da) and mass resolution (up to 2000) it is only used in proteomics with a combination of other analyzers. [63]

The TOF mass analyzer is based on the principle, that if ions with the same charge are accelerated by an electric field, they gain the same amount of kinetic energy. Their velocity is then determined by their mass, and subsequently so is the time it takes them to reach the detector. TOF analyzers work in a pulsed manner, where the accelerating field is perpendicular to the direction the ions are coming from. Based on flight times and the known parameters, a mass spectrum can be generated. The TOF analyzer has a virtually infinite mass range, high resolution (up to 10^5), and good mass accuracy, furthermore a very low scan time. In proteomics, it is usually used in combination with a quadrupole and an ESI source, or another TOF analyzer and a MALDI source. [64,65]

The ion trap analyzer comprises of two electrodes facing each other and a third ring electrode between them, creating a space where ions are trapped by static and oscillating electric fields. The FT-ICR is similar, but an additional uniform magnetic field is involved leading to the oscillation of ions. In an Orbitrap analyzer, ions are trapped oscillating between an outer (barrel) and an inner (spindle) electrode. Ion trap analyzers have a low mass range and resolution, but very high sensitivity; they are usually combined with an Orbitrap. The FT-ICR has the highest mass resolution, however, as it requires the use of superconductive magnets, its operation is very expensive. Orbitrap analyzers have very high resolution ($>10^6$), mass accuracy, and are capable of multistage fragmentation. They are widely used for proteomics applications combined with ion trap and quadrupole analyzers. [66–68]

1.3.3. Tandem Mass Spectrometry and Fragmentation Methods

Mass analyzers are either used by themselves (e.g., TOF-MS or FT-ICR-MS), or more frequently in tandem with each other (MS/MS), for example in a quadrupole time-of-flight setup (*Figure 2*).

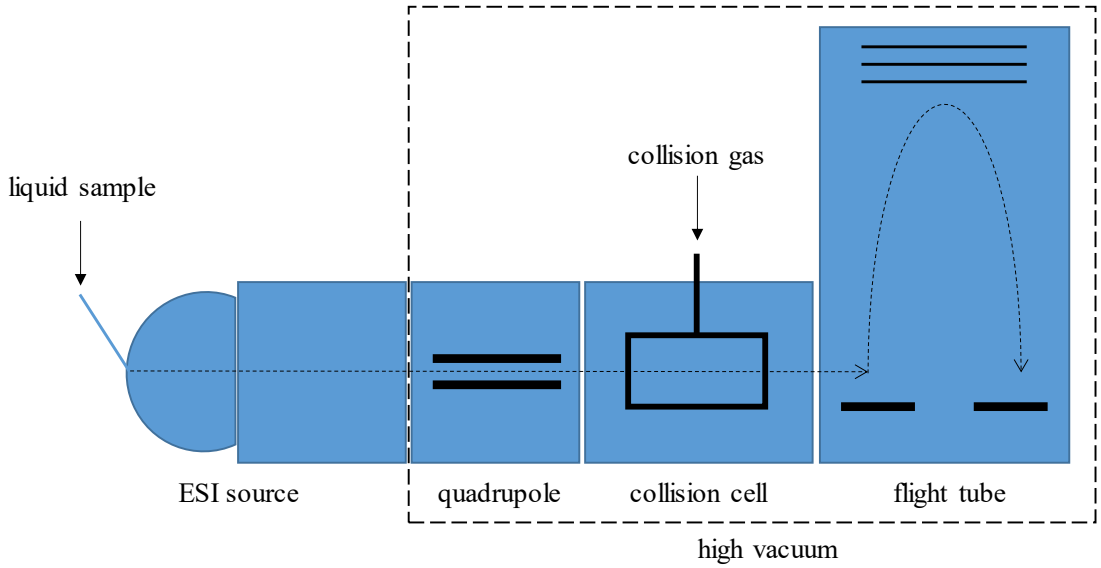

Figure 2: Schematic layout of a quadrupole time-of-flight (Q-TOF) mass spectrometer. The flight tube is shown in a reflectron arrangement—where the flight path is elongated with the use of ion optics—with an orthogonal accelerator and a detector. Figure made by the author using Microsoft PowerPoint.

Tandem MS can combine the strengths of two mass analyzers—tandem in space—or can be accomplished by a single ion trap analyzer—tandem in time; and allows the controlled fragmentation of ions inside the instrument. This enables the collection of more detailed information on the analytes based on their fragmentation characteristics. In proteomics, the main techniques are **collision-induced dissociation** (CID), and **electron transfer dissociation** (ETD). [69,70]

In CID, the precursor ions are accelerated to a certain energy, then allowed to collide with a neutral gas in the fragmentation cell. Inelastic collision occurs, when part of the ion's kinetic energy is converted into internal vibrational energy, which subsequently results in bond cleavage via the lowest energy pathways. Based on the acceleration energy, the precursor ion can undergo one, or multiple sequential fragmentation steps, which can be modulated based on the fragments of interest. Here, it is important to mention higher-energy collision dissociation (HCD), which is essentially the application of the CID method to Orbitrap instruments. [71–74]

The ETD method can be used to fragment multiply charged protonated molecular ions. It is based on ion-ion reactions, where a radical anion transfers an electron to the precursor cation subsequently forming an unstable radical, which undergoes fragmentation along the peptide backbone. The radical anions are generated from fluoranthene using chemical

ionization, then mixed with the precursor ions, usually in an ion trap type instrument. As opposed to CID (and HCD), in ETD the weaker bonds, e.g., PTMs do not undergo fragmentation, and especially large peptides, or even proteins can be dissociated as well. [71,73,75]

1.3.3.1. Fragmentation of Peptides and *N*-Glycopeptides

Fragmentation characteristics primarily depend on molecular attributes, the type of mass spectrometer (the dissociation method) used, and experimental conditions. Peptides and *N*-glycopeptides have inherently different molecular characteristics, which subsequently leads to different fragmentation behavior as well. Fragment ions derived from peptides and glycopeptides have their own nomenclature introduced by Roepstorff, and Domon-Costello (for carbohydrates in general), respectively (*Figure 3*).

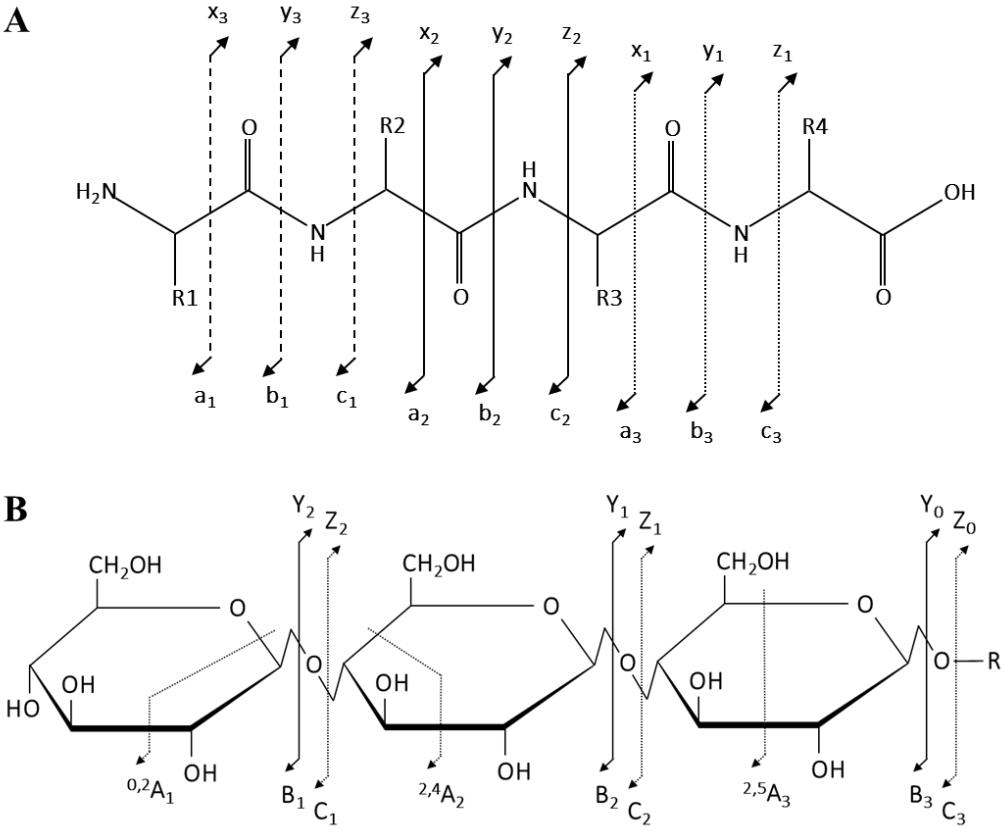

Figure 3: Fragmentation nomenclature for (**A**) peptides introduced by Roepstorff, and (**B**) for carbohydrates introduced by Domon-Costello. Figure made by the author using Microsoft PowerPoint.

The dissociation method used determines which fragment ions are produced. In CID (and HCD), peptides primarily produce ***b* and *y* type** ions, while in ETD ***c* and *z* ions** are produced. As in CID (and HCD) the weaker bonds fragment first, glycopeptides usually

fragment by their glycosidic bonds, while the peptide backbone remains intact. This only yields information on the glycan structure in the form of ***B* and *Y* type ions**, which usually means the peptide sequence cannot be determined with high confidence. In ETD on the other hand, the peptide backbone will undergo fragmentation, which leads to fragments informative of the peptide sequence as well, however, the glycan structure usually remains intact. As for glycopeptides HCD and ETD yield complementary information, there are applications in Orbitrap MS where they are combined. The most prevalent technique is called EThcD, where glycopeptides first undergo ETD, then secondary HCD, which leads to more informative mass spectra. [76,77]

Experimental conditions can highly affect fragmentation as well. In CID, fragmentation energy is routinely modulated based on the *m/z* of the precursor for optimal fragmentation. At higher energies multi-step fragmentations can occur, where glycopeptides first lose their glycan then the peptide backbone undergoes fragmentation as well (this is preferred for peptides). At lower energies, however, single-step reactions dominate, which are more informative of the glycan structure. These two energies can be combined—fragmentation time is split up between them—which results in a **mixed-energy spectrum** that contains informative fragments of both the glycan structure and the peptide backbone. [78]

1.3.4. Liquid Chromatography Coupled Mass Spectrometry

In proteomics, mass spectrometry is usually coupled to **high performance liquid chromatography** (HPLC) with the help of an ESI ion source, which makes the analysis of highly complex biological samples possible. The separation of complex mixtures before MS/MS analysis is beneficial for several reasons, e.g., it lowers ion suppression as fewer analytes compete for charge in the ESI source increasing sensitivity and dynamic range; makes high-throughput identification and quantitation possible; and reduces the number of problematic “chimeric” mass spectra—MS/MS spectra with fragments from multiple precursor ions. [79,80]

For proteomics applications HPLC is usually used in the **nanoflow range** (nanoUHPLC), as it provides improved sensitivity compared to higher flow rates (and larger column diameters). The columns used for peptide separation are still packed columns primarily; however, monolithic columns, open-tubular, and 3D printed pillar array columns have emerged recently as alternatives. [81] As proteomics is a diverse field (including the

analysis of multiple PTMs and other specific applications) several stationary phases are routinely used. For peptides, reversed-phase (RP) columns (e.g., the most frequently used octadecyl carbon chain (C18)-bonded silica) are the most common, which separate peptides based on their hydrophobicity. [82,83] For glycopeptides, apart from RP columns, normal phase, HILIC, PGC, and ion-exchange (predominantly weak anion exchange) columns are also routinely used, as they offer different selectivity. [84,85]

1.3.5. Mass Spectrometry Data Acquisition

In tandem MS, there are multiple options for data acquisition strategies depending on the task at hand. When the analysis is focused on known analytes, targeted approaches are employed, e.g., selected or multiple reaction monitoring, [86] and parallel reaction monitoring. [87] These methods, however, are seldom used in proteomics. On the other hand, when there is no prior knowledge available of the sample (in the majority of cases), discovery or untargeted approaches are used, usually **data dependent** (DDA) or **data independent** (DIA) **acquisition**.

The main differences between the two approaches are that the use of cycle times (time slots allocated for MS measurements during acquisition), and isolation widths (the mass range selected for fragmentation) are different. In DDA, first an MS spectrum is acquired, then based on this, a set maximum number of specific ions are selected for fragmentation (precursor ions). [88] Contrary, in DIA MS/MS spectra are acquired systematically, not based on information from precursor spectra and the entirety of a wider mass range is fragmented parallel. [33,89,90] There are several DIA methodologies available, including sequential window acquisition of all theoretical mass spectra (SWATH), [91] scanning quadrupole DIA (SONAR), [92] and parallel accumulation-serial fragmentation combined with data independent acquisition (diaPASEF). [93]

The strength of DDA is the relatively easy interpretation of mass spectra, as (ideally) an MS/MS spectrum contains fragments from only one precursor. On the other hand, due to the semi-stochastic nature of the DDA algorithm, the reproducibility of measurements is rather low, and less abundant ions are often overlooked. Compared to DDA, DIA methods provide more precise and reproducible quantitation, and more protein identifications. On the other hand, mass spectra generated are highly complex and hard to interpret, which can pose challenges especially for the analysis of PTMs. [34,94]

1.3.6. Protein and Glycoprotein Identification

The identification of proteins—in data-dependent bottom-up methodology—is an automated multi-step process called **database search**, that can be performed by several software employing different algorithms. The main steps are usually i) the generation of peptides and peptide spectra in silico from an appropriate protein database ii) an automatized search of all MS/MS spectra against the previously generated database resulting in peptide-spectrum matches (PSMs) iii) identification of proteins or protein groups (sequence homology-related ambiguity is a common feature) based on their corresponding peptides. Software settings can have a great impact on the results, the most important (and cross-platform) ones are the instrument parameters (fragmentation type, mass accuracy), the chemical modifications introduced at sample preparation, the protein database used, and the false discovery rate (FDR). [95–97]

For *N*-glycoproteins, software solutions and methodologies are not as well established as for proteins. There are a lot of options available, however, there are some key questions not completely resolved yet, e.g., methods for calculating glycopeptide FDR. [98,99] Search engines use similar methods as for proteins but with the different fragmentation characteristics of glycopeptides considered. Furthermore, some of them employ glycopeptide-specific approaches as well, e.g., pre-filtering mass spectra for glycopeptide specific fragment ions, or accounting for the estimated retention time differences between glycoforms. [85,100,101]

1.3.7. Quantitation in Proteomics

There are several methods suitable for quantitation in proteomics, but it is usually based on the measurement of chromatographic peak areas. This metric can be affected by several factors apart from analyte concentrations, e.g., instrument parameters, matrix effects, and differences in analyte ionization efficiencies. Because of these, quantitation is not straightforward; there are two fundamental approaches in high-throughput proteomics, relative and absolute quantitation. In **relative quantitation**, samples are compared to each other, and the results will only reflect the ratio of the analyte for these samples (it is not, or rarely comparable with other separate experiments). This can either be achieved with the use of stable isotope labeling (chemical, proteolytic, or metabolic labeling) or without it (label-free quantitation). Contrary, in **absolute quantitation** exact protein amounts (e.g., the concentration of a target protein in a cell lysate) can be

determined for individual samples with the help of stable isotope labeled internal standards. Labeling-based strategies are more accurate and reproducible than **label-free quantitation** (LFQ), but are more expensive, labor-intensive, and not applicable in all circumstances (e.g., for the analysis of human tissue samples). [102–104]

1.4. Proteomics in Cancer Research

Proteomics in cancer research in general aims at the better understanding of molecular changes with tumorigenesis and progression, and subsequently the discovery of clinically relevant molecules, e.g., diagnostic and prognostic markers, or therapeutic targets. The focus, however, can be very different for specific cancers, as they all have different incidence, mortality, methods for screening, diagnosis, and therapeutic options, creating unique problems and challenges to be solved.

1.4.1. Prostate Cancer Proteomics

Prostate cancer research—including proteomics—is currently focused on two main topics: finding novel diagnostic markers to replace the PSA blood test, and prognostic markers that accurately predict aggressive PCa.

Replacing, or improving the accuracy of the PSA blood test has been long on the forefront of PCa proteomics research, as it has low specificity towards PCa causing overdiagnosis and overtreatment. For this reason, there is significant effort directed towards finding novel biomarkers to replace the PSA blood test. Usually, blood, prostate tissue, urine, seminal plasma, and prostate-specific exosomes are analyzed using shotgun or targeted proteomics. In recent years, the focus shifted from finding a single marker to a panel for increased accuracy. There are several promising putative markers—including several different protein panels—currently under investigation [105], however, none of them have been implemented into clinical practice in a widespread manner yet. [106–111]

PSA *N*-glycosylation-specific changes have also been extensively investigated as a potential PCa marker. [112,113] There are significant challenges in determining PCa aggressiveness: tumor heterogeneity, biopsy-sampling error, and possible variation in biopsy interpretation, which contribute to significant overtreatment of patients with low-risk PCa. This makes the discovery of novel prognostic markers especially important. [108] The methodologies and sample types analyzed are the same as for diagnostic markers, and—although not as many—several protein panels are currently under different stages of investigation. [105,110,111]

Although not as central a research field as the two mentioned above, improving therapeutic options for advanced PCa (the treatment of low-risk PCa is relatively well-established) along with drug resistance and therapeutic response markers is an important and significant part of PCa research as well. [114,115]

1.4.2. Lung Cancer Proteomics

Currently, lung cancer (LC) proteomics studies are mostly focused on discovering novel diagnostic, prognostic, and predictive molecular markers [116,117]; and potential therapeutic targets [118–120] to improve upon currently available options. The analyzed sample types include cell lines, body fluids (blood, urine, saliva etc.,), and fresh frozen (FF) or formalin-fixed paraffin embedded (FFPE) tissue. [116]

There is an abundance of cell lines established from SCLC and NSCLC cells, and they are widely used by the proteomics community for LC research. The main advantages of their use are the easy availability, unlimited sample amounts, and wide selection (more than for other epithelial cancers). On the other hand, there are serious disadvantages: genetic instability along with selective growth of subpopulations during long term passage, the absence of interactions with non-tumor components, and inherently the lack of biological replicates. [121] For these reasons the use of cell lines is usually limited to basic research.

The study of biofluids in LC proteomics is focused on two clinically relevant areas: early cancer diagnosis; and prognostic, and predictive biomarkers. LC screening is primarily done by imaging techniques (chest radiography, or low-dose computed tomography), however, the results of these tests can be inconclusive because of the presence of indeterminate pulmonary nodules, where biopsy might not be an option. [116] In these cases, early cancer markers, such as proteins or specific proteoforms in biofluids could aid the diagnostic process. There is a number of early-cancer protein markers (protein panels) currently in the clinical validation phase. [116] After LC is confirmed, it needs to be classified, staged, and therapeutic decisions need to be made. Here, the main benefit of using biofluid markers is the option for constant monitoring, as opposed to biopsies they are minimally invasive, and can be more frequently taken. [122] Currently, the primary sample types studied are blood, sputum, bronchial lavage or aspirate samples, urine, and saliva. [116]

The proteomic analysis of tissue samples is central to understanding LC biology, as it enables the selective study of certain subtypes of the disease in an in vivo setting, and also intratumoral heterogeneity. [123] Furthermore, there are multiple tissue-based markers currently in clinical use, several of them proteins: immunohistochemical markers (e.g., thyroid transcription factor 1, napsin A, for AC; p40, and cytokeratin 5/6 for SqCC), genomic markers (e.g., epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), or Kirsten rat sarcoma viral oncogene homolog (KRAS)), and immunotherapy markers (e.g., cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4), and programmed death-ligand 1 receptor (PD-1)). [124]

1.5. Special Considerations in the Analysis of FFPE Tissue Samples

In proteomics, there is a wide variety of sample matrices that require special treatment before analysis, e.g., depletion of plasma samples. For FFPE tissue samples this is no different. FFPE tissue samples are most frequently analyzed in the form of thin slices mounted to microscopy slides. There are two major aspects to consider before proteomics analysis. First, the removal of paraffin, and second, breaking the protein cross-links that are formed during formalin fixation—called antigen retrieval. [125,126]

Deparaffinization is most frequently achieved by washing the tissue with xylene; however, alternative xylene-free methods have been proposed as well. [127,128] Antigen-retrieval (AR) is a heat-induced process, that requires submerging the slides in a heated buffer solution for an extended period of time, while the rehydration of the tissue also happens. There are several buffer solutions used, e.g., Tris-EDTA, or sodium citrate. [129,130] As the deparaffinization and AR is usually a mechanically intensive process, tissue slides are usually baked at medium-high temperatures beforehand to ensure proper fixation to the microscope slide.

FFPE tissue blocks can be stored for extended periods of time; however, there has been concern about the effect of formalin fixation on the analysis of proteins. There have been several comparisons between FFPE and fresh-frozen tissue, and although there are differences, e.g., there are fewer lysin residues in FFPE tissue, there is mounting evidence, that FFPE tissue is a feasible alternative to fresh-frozen tissue in retrospective and prospective studies. Additionally, for the comparability of FFPE tissue specimens a crucial aspect to consider is that the protocol (especially the time) of formalin fixation needs to be standardized. [131,132]

2. Objectives

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

The first project—the characterization of prostate TMA samples [133]—was part of the "Mass spectrometry based identification of prostate cancer biomarkers from tissue microarrays" project funded by the National Research, Development and Innovation Office (OTKA PD 121187). The aim was the integration of previous research and method development, and its application to the analysis of prostate TMAs to identify molecular alterations with potential diagnostic value. The most important prerequisites were the successful application of on-surface tryptic digestion to prostate TMAs that contain very limited sample amounts (approximately 1 µg in total) in a proteomic pilot study comparing healthy and cancerous samples [134], and the optimization and application of a glycopeptide enrichment method—acetone precipitation—to similar sample amounts. [135] The key objectives of the project were: successful combination of on-surface digestion and acetone precipitation to obtain proteomic and *N*-glycoproteomic information in parallel from the same TMA sample, and the investigation of different grades of PCa for a better understanding of disease progression.

The second project—the proteomic analysis of lung tissue sections [136]—was part of the "Identification of predictive biomarkers for the prognosis of targeted biological therapy in lung cancer" project funded by the National Research, Development and Innovation Office (OTKA FK 131603). Here, the objective was to identify dysregulated biological processes by analyzing and comparing the four main types of LC (small cell lung cancer, adenocarcinoma, squamous cell carcinoma, and large cell carcinoma) with their adjacent normal regions and each other by proteomics using on-surface tryptic digestion of FFPE tissue, enabling a more precise diagnosis and classification of lung cancer.

3. Methods

3.1. Materials

All chemicals used were HPLC-MS grade except for acetone (HPLC grade). Acetonitrile (ACN), water, acetone, formic acid (FA), ammonium-bicarbonate, citric acid, trisodium citrate, and heptafluorobutyric acid (HFBA) were purchased from Merck. Trifluoroacetic acid (TFA), dithiothreitol, and iodoacetamide were obtained from Thermo Scientific. Methanol was purchased from VWR International, RapiGest surfactant was obtained from Waters, trypsin and LysC-trypsin were purchased from Promega.

3.2. Patient and Sample Information – Prostate Tissue

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

3.3. Patient and Sample Information – Lung Tissue

Formalin-fixed paraffin embedded AC, SqCC, LCC, SCLC (n = 10, 9, 10, 9 respectively) and tumor-adjacent normal tissue sections (n = 8, 8, 8, 9 respectively) were analyzed. The work was approved by the Medical Research Council (TUKEB permit number: IV/2567-4/2020/EKU).

Inclusion criteria for patients were the following: fresh LC cases with resection specimens, while also keeping in mind that histological groups should have similar sizes (the aim was 10 patients in each group). FFPE tissue sections with a thickness of 10 μm were obtained from the departmental archive of the Department of Pathology, University of Pécs, Hungary. Summarized information on the samples is provided in *Table 1*.

Table 1: Summary of patient and sample information for FFPE lung tissue specimens.

Sample characteristics	No. of Patients	No. of Tumor Samples	No. of Tumor Adjacent Samples
Total No.	40	38	33
Age	66 (54–79)		
Gender			
Male	22	21	19
Female	18	17	14
Histology			
Adenocarcinoma	10	10	8
Large cell carcinoma	7 (10)	7 (10)	5 (8)
Large cell neuroendocrine carcinoma	3	3	3
Small cell lung carcinoma	10	9	9
Squamous cell carcinoma	10	9	8

3.4. Sample Preparation

3.4.1. On-Surface Digestion

Tissue sections were baked at 60 °C for 2 h to prevent tissue detachment. Next, deparaffinization was carried out by sequentially incubating the slides in xylene for 2 × 3 min, in ethanol for 2 × 5 min, in 90:10 v/v% ethanol:water for 3 min, in 70:30 v/v% ethanol:water for 3 min, in 10 mM NH_4HCO_3 (ABC) for 5 min, and, finally, in water for 1 min. After dewaxing, heat-induced antigen retrieval was performed (for prostate TMAs: 20 mM Tris–HCl, pH = 9.0 for 30 min at 90 °C; for lung tissue samples: 95 mM trisodium citrate + 21 mM citric acid in water, pH = 6 for 30 min at 80–85 °C).

Following the preparation steps, digestion was carried out: for prostate TMAs the whole surface area, for lung tissue samples on specific tissue regions based on characterization by a pathologist. Each digestion step involved pipetting the reagents on the designated surface—for prostate TMAs 1 µL, for lung tissue 3 µL. The proteins were reduced using 0.1% RapiGest and 5 mM dithiothreitol in 20% glycerol for 20 min at 55 °C, then alkylated using 10 mM iodoacetamide in 25 mM ABC buffer and 20% glycerol for 20 min at room temperature in the dark. The digestion was performed cyclically, each one lasting for 40 min at 37 °C in a humidified box with 5 cycles in total. In the first two cycles, LysC-trypsin mixture was added in ca. 1:25 ratio in 50 mM ABC and 20% glycerol. Subsequently, in the last three cycles, trypsin was added in a 1:5 ratio in 50 mM ABC, and 20% glycerol. After digestion, the extraction of the protein digest was carried out by pipetting 10% acetic acid extraction solvent (1 µL for TMAs and 3 µL for lung

tissue) five times on the digested spots. Peptide extracts were then dried down and stored at −20 °C until further usage.

3.4.2. Reversed-Phase Purification

C18 spin columns were used for desalting and clean-up. After the column was conditioned, washed, and equilibrated, the sample was loaded onto the column in 0.1% FA for TMA samples, and 0.1% HFBA for lung tissue samples. The elution was performed with 30:70 v/v% water:ACN. After the elution, the samples were dried down and stored at −20 °C until further usage.

3.4.3. Glycopeptide Enrichment for Prostate TMAs

After purification samples were reconstituted in 15 µL 1% FA and 150 µL ice-cold acetone was added and the solution was stored at −20 °C overnight. Then the samples were centrifuged at 13,000 g for 10 min, then the supernatants were removed, dried down, and stored at −20 °C. The pellet fractions were also dried down, then resuspended in 10 µL of injection solvent and subsequently stored in the autosampler unit until analysis.

3.5. nanoUHPLC-MS(MS) Analysis

Samples were analyzed using a Maxis II QTOF instrument equipped with a CaptiveSpray nanoBooster ion source coupled to a Dionex UltiMate 3000 RSLCnano system. Peptides were separated on an Acquity M-Class BEH130 C18 analytical column (1.7 µm, 75 µm × 250 mm) using gradient elution (isocratic hold at 4% for 11 min, then elevating B solvent content to 25% in 75 min, and to 40% in 15 min) following trapping on an Acclaim PepMap100 C18 (5 µm, 100 µm × 20 mm) trap column. Solvent A consisted of 0.1% FA in water, Solvent B was 0.1% FA in ACN, sample loading buffer was 0.1% TFA for prostate TMA samples, and 0.1% TFA + 0.01% HFBA for lung tissue samples. For proteomics MS analysis, DDA measurements were performed. The cycle time was set at 2.5 s, with a dynamic MS/MS exclusion of the same precursor ion for 2 min, or if its intensity was at least 3 times larger than before. Preferred charge states were set between +2 and +5. MS spectra were acquired at 3 Hz in the 150–2200 m/z range, while MS/MS spectra were at 4 or 16 Hz, depending on the intensity of the precursor. Internal calibration was performed by infusing sodium formate and raw data was recalibrated using the Compass DataAnalysis software 4.3. For glycoproteomics MS/MS measurements, the experimental settings were similar, except for collision energies. Mixed energy spectra were collected at 100% collision energy for 80% of the cycle time

and 50% collision energy for 20% of the cycle time. For single-stage MS measurements, spectra were recorded over the mass range of 300–3000 m/z at 1 Hz.

3.6. Software Used

Software used for PCa proteomics and glycoproteomics: MASCOT, [137] MaxQuant 1.6.17, [138] Perseus 1.6.5.0, [139] R 3.6.1., [140] RStudio 1.2.5001, [141] Byonic 3.8, GlycoPattern 4.7_b30, [142] STRING web-application. [143]

Software used for LC proteomics: Byonic 3.8, MaxQuant 1.6.17, R 3.6.1, RStudio 1.2.5001, GSEA 4.2.3, [144] Cytoscape 3.9.1. [145]

3.7. Data Analysis

3.7.1. Prostate Cancer Proteomics and Glycoproteomics

Protein quantitation was performed by MaxQuant on a focused Homo Sapiens database made by combining MASCOT search results from all MS/MS analyses. The MaxQuant output was then loaded into Perseus, where proteins found in less than 60% of each sample group were removed. Subsequently, missing values were imputed from a normal distribution with the default settings for the whole matrix (down shift of 1.8 and width of 0.3). Statistical analysis was then performed, using two-sample tests (Student's t-test), multiple-sample tests (ANOVA), and post-hoc tests (Tukey's HSD). Data visualizations were done in RStudio using the ggplot2 library. In the glycoproteomics analysis, glycosites were identified from the HPLC–MS/MS analysis of pooled pellet samples using Byonic with a |LogProb| value of at least 2. The same HPLC–MS/MS experiments were used to identify the composition of various glycans at these glycosylation sites. GlycoPattern was then used to quantify the glycopeptides based on single-stage nanoUHPLC-MS corresponding to the linear combinations of the glycosites and glycans previously identified. The software identified the glycopeptides according to their exact mass, retention time, and isotope cluster distribution, then performed label-free quantitation. Pre-processing and statistical analysis were then carried out using R in RStudio. The data were first submitted to outlier filtering, where identifications with a retention time outside of the first quartile−1.5 IQR (inter-quartile range) to the third quartile+1.5 IQR range were thrown out. Then, through sequential filtering steps, any data points with an area under curve (AUC) less than 1000, glycopeptides identified in less than 5 samples, and samples with less than 10 glycopeptides identified were removed. Subsequently, the data were normalized using Quotient Total Area Normalization

followed by log transformation. The degree of fucosylation and sialylation were then calculated for every glycosite. Statistical analysis was carried out in Perseus similarly to the proteomics dataset. Functional Enrichment of proteins was performed in STRING for Gene Ontology (GO) Terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways. The minimum required interaction score was set to the highest confidence (0.900).

3.7.2. Lung Cancer Proteomics

Protein quantitation was performed in MaxQuant on a focused Homo Sapiens database, made from merging Byonic search results from all MS/MS analyses. Subsequent data analysis steps were performed in R using RStudio. After quantitation, the data was filtered based on the number of missing values in each of the eight sample groups: only proteins found in at least 2/3 of all samples in at least one sample group were kept for further analysis. Missing values were then imputed in a group-wise manner according to the following: if the protein in question was found in less than 2/3 of all samples in the group, it was imputed with the sample 5-percentile, while if it was found in at least 2/3 of all samples in the group, it was imputed according to the kNN algorithm (VIM package, k = 15, similarity based on Euclidean distances, default settings used). Following imputation, statistical analysis was carried out. For the different comparisons (multiple and two-group comparisons) within a group, normality and equality of variances were tested (Shapiro-Wilk tests and Levene tests, respectively). For multiple group comparisons ANOVA, Welch-ANOVA, and Kruskal-Wallis tests, for two-group comparisons, Student's t-tests, Welch t-tests, and Wilcoxon rank sum tests were performed based on the outcome of the normality and variance equality tests. FDRs were controlled for all two-group and all multiple-group comparisons separately using the Benjamini-Hochberg method at 5%. Plots were made using the ggplot2, ggpattern, gplots, and nVennR packages. The principal component analysis (PCA) was performed using the prcomp function (using variable scaling and default settings), hierarchical clustering was performed using the heatmap.2 function (using Ward's clustering method "ward.D2" from the hclust function). The identification of enriched gene sets was performed with the GSEA software using the GSEAPreranked function based on effect sizes (Cohen's d) calculated for the four types of LC separately (adjacent vs. tumorous tissue). For gene sets, the Hallmark, KEGG, and GO databases available in GSEA were used.

4. Results

The *Results* section is divided into two parts based on the tissue type—and subsequently cancer type—studied. This explicit separation is not only necessary to prevent confusion of similar but unrelated results from the two separate projects, but also helps avoid their direct comparison, which is undesirable (see *5. Discussion* for details).

4.1. Prostate Cancer Proteomics and *N*-Glycosylation

During the project aiming at the proteomic and *N*-glycoproteomic characterization of cancerous and normal prostate tissue, 95 TMA biopsy samples were analyzed. Among these, there were 9 Grade 1 (G1), 16 Grade 2 (G2), 24 Grade 3 (G3), and 46 normal (N). The sample preparation consisted of on-surface tryptic digestion followed by C18 SPE cleanup and acetone precipitation for glycopeptide enrichment. After precipitation, the glycopeptide-enriched pellet fraction and the supernatant fraction containing non-glycosylated peptides were analyzed separately (see *Methods* for experimental details).

This section is divided into three major parts: i) the molecular differences between healthy and cancerous prostate tissue; ii) the molecular changes with PCa grade progression, and differences between distinct grades and healthy tissue; iii) and the biological processes altered in PCa. While the first two sections are based on data from both the proteomics and glycoproteomics datasets, the third one is based on proteomics data only. Before describing the results of the three sections, a general characterization of the two datasets (proteomics and glycoproteomics) is provided.

Proteins were quantified using MaxQuant, 653 proteins were quantified altogether in the 95 supernatant samples analyzed. From these, proteins that were found in less than 60% of any of the sample groups were excluded. Missing values were then imputed. *N*-linked glycopeptides were quantified using the GlycoPattern software using glycan and glycosite libraries constructed following database searches using the Byonic software. Results were then filtered (see *Methods* for details on data analysis). Altogether 145 glycopeptides were quantified in 95 samples with high confidence, corresponding to 22 glycoproteins with 29 glycosites and 53 different glycans.

Protein glycosylation can be characterized by listing all the identified glycopeptides, but usually, multiple metrics are used instead. Here sialylation, fucosylation, galactosylation, and glycan type ratio are used. These simplify data interpretation and carry important biological information as well, as they are connected to various steps of *N*-glycan

biosynthesis. Sialylation and galactosylation are calculated as the proportion of glycan antennae containing the respective monosaccharide versus all antennae. Fucosylation is calculated as the ratio of fucose-containing glycans to all glycans. The glycan type ratio is the proportion of glycans classified as either complex, oligomannose, hybrid or paucimannose (uncommon, truncated glycans smaller than the core structure).

Over 75% of the identified glycopeptides carried complex-type glycans. More than half of these structures were biantennary, while tri- and tetra-antennary types and unmatured structures were also present. The average antenna sialylation was 20.1% across all samples, while 28.7% of antenna containing structures held at least one sialic acid. The average fucosylation was 37.8% across all samples. All 29 glycosites identified carried several different glycans, and showed considerable diversity regarding glycan type, number of antennae, galactosylation, fucosylation, and sialylation. To reveal changes specific to the distinct glycosites, metrics were calculated for them individually as well.

4.1.1. Differences between Normal and Cancerous Tissue

To investigate differences between normal and cancerous tissue, t-tests were used with a 0.05 false discovery rate. In the proteomics dataset 123 proteins were found to be differentially expressed, this included 72 proteins overexpressed and 51 proteins underexpressed in PCa. Among these, 14 showed a fold-change over 2, while 27 displayed a fold-change under 0.5. In the glycoproteomics dataset, 7 glycopeptides were found with significantly different abundances between the normal and PCa groups, each carrying biantennary, fucosylated complex-type glycans with different levels of galactosylation and sialylation. In five cases, glycopeptide abundance was lower in PCa tissue: four glycoforms of immunoglobulin gamma-1 heavy chain (IGG1) N299, and one glycoform of prothrombin (THRB) N121. In the other two cases, glycopeptide abundance was higher in PCa: one glycoform of microfibril-associated glycoprotein 4 (MFAP4) N137 and one glycoform of biglycan (PGS1) N270. Differences were also detected between normal and PCa tissues when comparing the levels of sialylation, fucosylation, and galactosylation at distinct glycosites. The differences in glycosite-specific sialylation, fucosylation, and galactosylation are summarized in *Figure 4*.

DOI:10.14753/SE.2023.2877

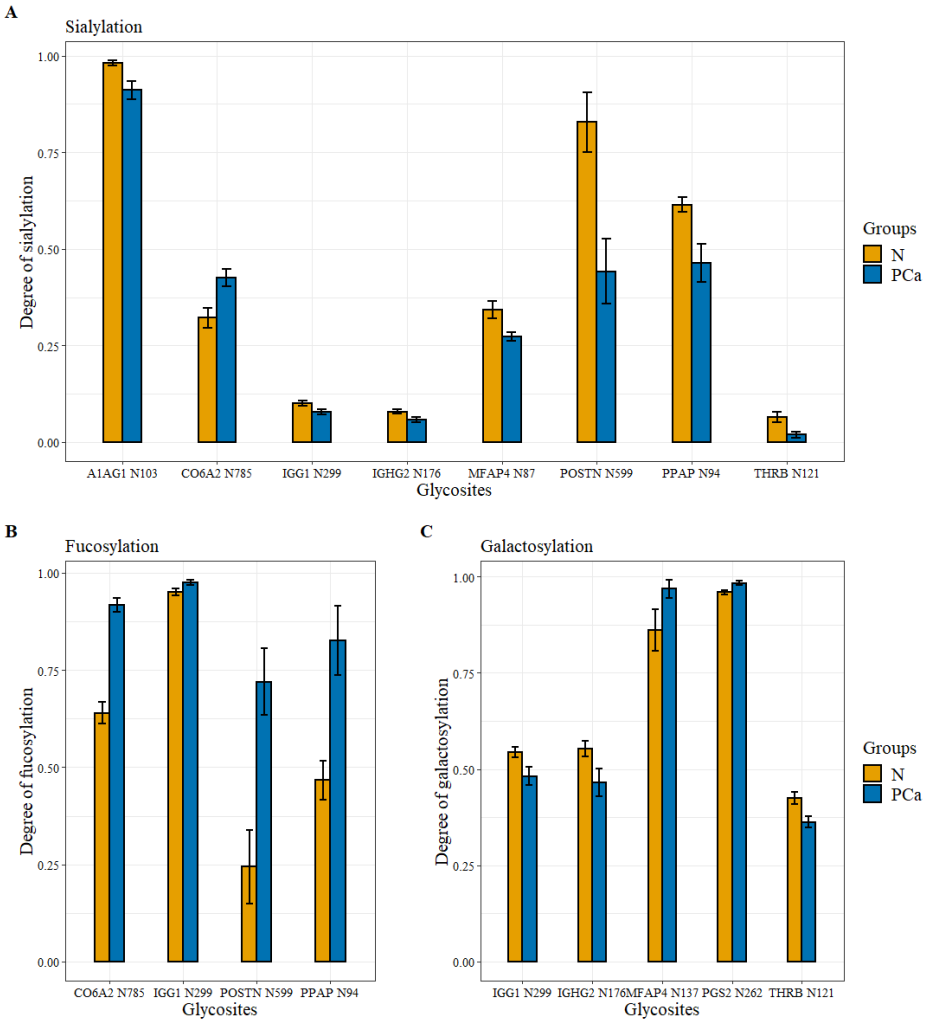

Figure 4: Glycosite-specific alterations in sialylation (**A**), fucosylation (**B**), and galactosylation (**C**) between healthy and PCa tissues (with standard error displayed). Figure made by the author with ggplot2 in R.

All but one of the eight differentially sialylated glycosites showed decreased sialylation in PCa (*Figure 4A*). The difference was below 10% for most of these, except for periostin (POSTN) N599 and prostatic acid phosphatase (PPAP) N94 with a 38.6% and 15.1% decrease respectively, and collagen alpha-2(VI) chain (CO6A2) N785 with a 10.3% increase in sialylation. Although only a 4.6% difference, THRB N121 showed the greatest relative change with a degree of sialylation almost 3.5 times lower in PCa than in normal tissues. As opposed to this, all four differentially fucosylated glycosites were overexpressed in PCa with the biggest differences on N785 of CO6A2, POSTN N599, and PPAP N94 with a 27.7%, 47.6%, and 35.9% increase in fucosylation, respectively (*Figure 4B*). The differences in galactosylation levels found on five glycosites (*Figure 4C*) were much smaller than changes in fucosylation or sialylation, the two major ones

being the increase of galactosylation at MFAP4 N137 by 10.6% and the decrease of galactosylation of immunoglobulin heavy constant gamma 2 (IGHG2) N176 by 8.7% in case of cancerous samples. Interestingly, while the directions of the changes in fucosylation were consistent (fucosylation increased in PCa), in the case of sialylation and galactosylation they were not (*Figure 4A-C*).

4.1.2. Differences with Prostate Cancer Grade Progression

To uncover molecular alterations among pathological grades and normal tissue, ANOVA was used with FDR controlled at 0.05. In the proteomics dataset, 75 proteins were identified with significant changes among the various PCa grades and normal tissue. Hierarchical clustering based on Spearman's correlation revealed two distinct groups among these proteins: 40 proteins were mostly upregulated (*Figure 5A*), while 35 downregulated (*Figure 5B*) with cancer progression.

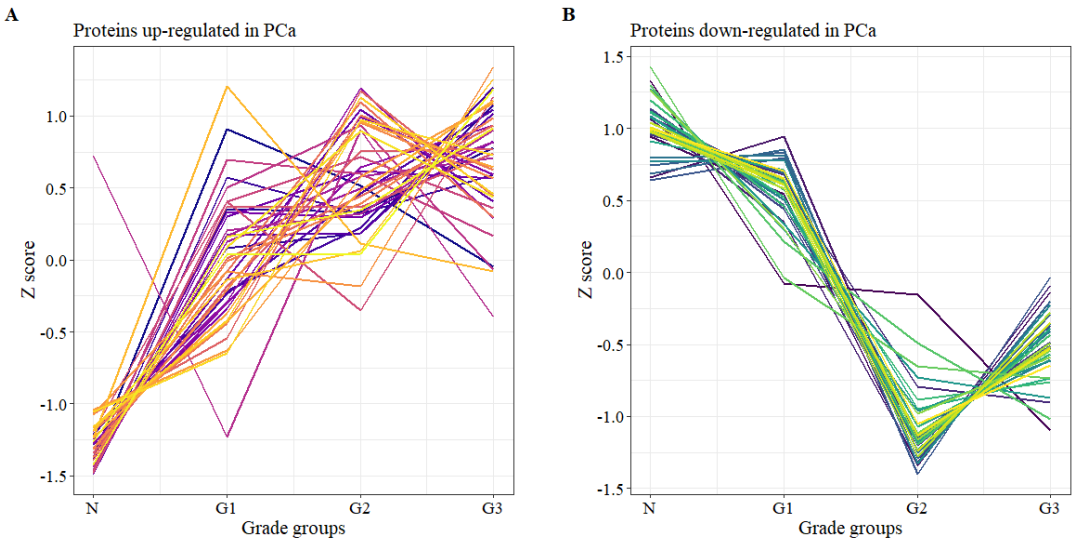

Figure 5: Proteins with changes among different grades of PCa and healthy tissue divided into two sub-groups based on hierarchical clustering: up-regulated (**A**) and down-regulated (**B**). Figure made by the author with ggplot2 in R.

Afterwards, a post-hoc test was performed on the 75 ANOVA significant proteins (Tukey's HSD test). This revealed that most of the proteins were differentially expressed between the normal and the two high-grade groups (G2 & G3), while there were only 3 such proteins between G2 and G3, 8 proteins between G1 and G3 and 14 between normal and G1 groups. Furthermore, many of them (more than 85%) showed differential expression in not only one but multiple group comparisons.

In the glycoproteomics dataset, ANOVA and the following post-hoc test (Tukey's HSD) revealed 4 glycopeptides with significantly different abundances among different grades and healthy tissue. Three of them correspond to the same glycosite N299 of IGG1 and carry biantennary complex-type glycans. In all three cases, the significant differences were between the Normal–G2 and Normal–G3 groups, and the observed trends were similar (average correlation coefficient of 0.98). The overall amount of IGG1 glycopeptides did not change significantly with PCa progression. The fourth glycopeptide corresponds to glycosite N137 of MFAP4 and also carries a biantennary complex-type glycan. In this case, the significant difference is between the Normal–G1 groups.

Furthermore, regarding glycosites, the degree of fucosylation on CO6A2 N785 was found to be different between the three Grade groups and Normal tissue. Interestingly, fucosylation shows a monotonic increase until G2 then decreases in G3 (*Figure 6A*). This tendency is very different from the changes in protein expression levels of the 3 identified CO6 subunits A1, A2, and A3 (*Figure 6B*).

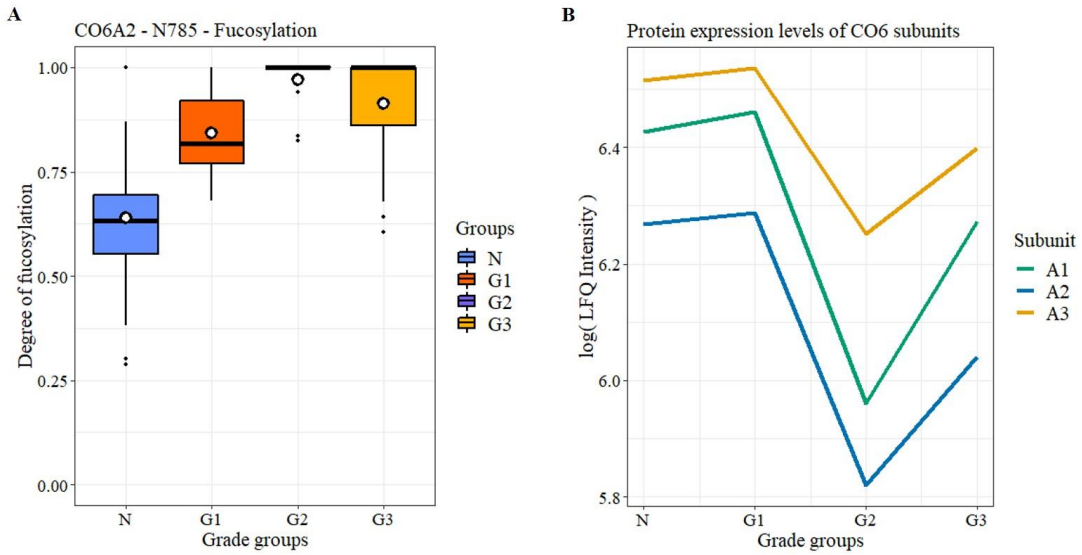

Figure 6: Changes in the fucosylation of CO6A2 (**A**) and in the protein expression of different CO6 subunits (**B**) between different Grades of PCa and healthy tissues. Figure made by the author with ggplot2 in R.

4.1.3. Biological Processes Altered in Prostate Cancer

Following the identification of proteins with statistically significant changes, functional enrichment analysis was performed in STRING for GO and KEGG terms, separately for the proteins up- and downregulated in PCa. Functional enrichment provides information on whether certain biological terms are overrepresented (enriched) in the list of proteins

provided. The most important terms from the resulting Protein–Protein Interaction (PPI) Networks were identified based on the Number of Genes (the number of term-associated genes identified), Strength (a metric that describes how large the enrichment effect is), and FDR values (describes the confidence in the results) and are summarized in *Table 2*.

Table 2: The most important terms in the PPI Networks from the STRING analysis including important metrics like number of genes (proteins) involved, strength of enrichment, and FDR.

Protein Pathways underexpressed in PCa

Database	Description	Number of Genes	Strength	FDR
KEGG	Focal adhesion	12	1.39	6.62×10^{-12}
KEGG	ECM-receptor interaction	6	1.47	2.49×10^{-6}
KEGG	Vascular smooth muscle contraction	6	1.30	1.44×10^{-5}
GO BP	muscle contraction	13	1.33	6.53×10^{-11}
GO BP	actin filament-based process	14	1.05	7.47×10^{-9}
GO BP	cell junction assembly	9	1.43	2.94×10^{-8}
GO BP	cellular component organization	34	0.42	1.30×10^{-7}
GO BP	actin cytoskeleton organization	12	1.06	1.30×10^{-7}
GO BP	extracellular matrix organization	10	1.13	6.01×10^{-7}
GO BP	cytoskeleton organization	14	0.77	7.79×10^{-6}
GO BP	supramolecular fiber organization	9	0.97	4.79×10^{-5}
GO BP	platelet degranulation	6	1.27	8.66×10^{-5}
GO BP	cell adhesion	12	0.75	8.66×10^{-5}
GO BP	regulated exocytosis	11	0.80	8.66×10^{-5}

Protein Pathways overexpressed in PCa

Database	Description	Number of Genes	Strength	FDR
KEGG	Ribosome	8	1.22	4.25×10^{-6}
KEGG	Spliceosome	8	1.22	4.25×10^{-6}
GO BP	mRNA metabolic process	23	0.97	4.28×10^{-13}
GO BP	RNA splicing, via transesterification reactions	15	1.15	2.64×10^{-10}
GO BP	protein localization to endoplasmic reticulum	9	1.30	1.76×10^{-7}
GO BP	metabolic process	60	0.23	1.92×10^{-7}
GO BP	SRP-dependent cotranslational protein targeting to membrane	8	1.37	3.02×10^{-7}
GO BP	translational initiation	9	1.24	3.57×10^{-7}
GO BP	protein localization	25	0.54	8.45×10^{-7}
GO BP	negative regulation of gene expression	23	0.57	8.87×10^{-7}
GO BP	regulation of gene expression	38	0.36	2.04×10^{-6}
GO BP	cellular response to cytokine stimulus	17	0.69	2.08×10^{-6}
GO BP	localization	41	0.33	2.51×10^{-6}
GO BP	translation	11	0.92	3.63×10^{-6}

Most of the underexpressed proteins were associated with cellular component organization (34 out of 51), while the overexpressed proteins were predominantly affiliated with metabolic processes (60 out of 72).

4.2. Lung Cancer Proteomics

To characterize the proteomic differences between the different types of lung cancer (AC, SqCC, LCC, SCLC), 71 samples were analyzed. These were derived from FFPE tissue sections taken from individuals suffering from either of the four different types of LC. From the tissue sections, samples were taken from both the cancerous and the cancer-adjacent normal regions (in a few cases the latter was not available); 8–10 samples belong to each sample group (patient and sample information is summarized in Table 1 in *Methods*). Of the ten samples with large cell morphology, three were large cell neuroendocrine carcinoma (LCNEC) according to the diagnostic criteria of the latest WHO classification. LCNEC is part of the neuroendocrine carcinoma group (besides the more common small cell neuroendocrine carcinoma), but because of the small size of the LCNEC group, and the similarities in microscopical morphology, these three LCNECs were grouped with LCC samples.

During the initial protein identification step—used for the construction of a focused protein database—9316 different proteins were identified altogether by Byonic software, on average ca. 1600 from the individual samples, as peptides from 3 µL digestion areas on the tissue surface were extracted and analyzed. Using the focused database, 2917 proteins were quantified by MaxQuant using label-free quantitation, out of which 1345 were considered for statistical analysis after initial filtering (see *Methods* for details on data analysis). Principal component analysis (PCA) showed that there are considerable differences between the tumorous and the tumor-adjacent regions, especially when considering a single LC type at a time (*Figure 7*).

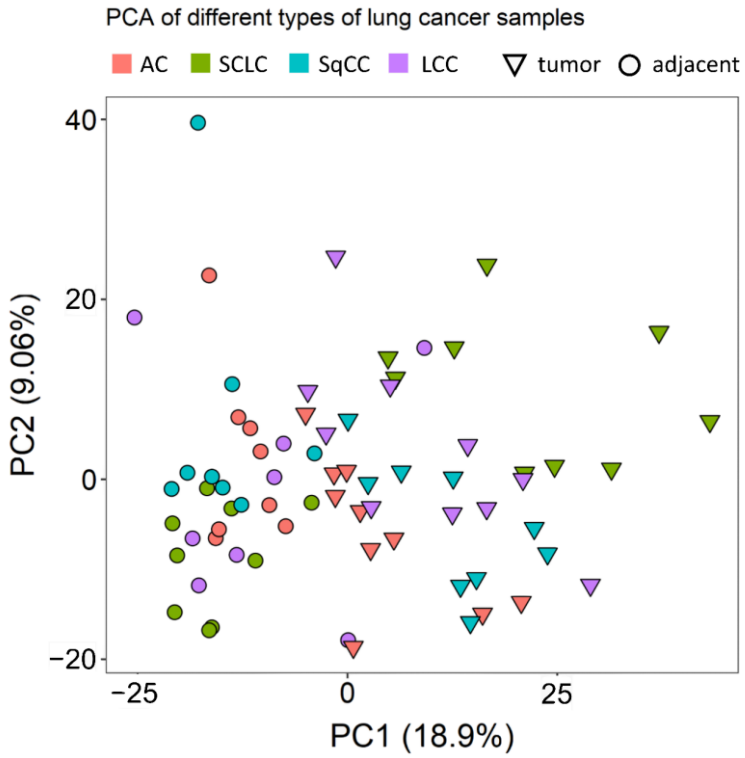

Figure 7: PCA of the different samples analyzed. Triangles and circles indicate tumor and tumor-adjacent samples, respectively. Different colors mark different LC types (red for AC, green for SCLC, turquoise for SqCC, and purple for LCC). Figure made by the author with ggplot2 in R.

Following the initial assessment, the data was further investigated to identify proteins that are differentially expressed between i) all adjacent and all tumor tissue, ii) adjacent and tumor tissue in each type, and iii) the different types of tumor tissue.

4.2.1. General Differences between Tumor and Adjacent Normal Tissue

To identify differences between adjacent and tumor tissue (33 and 38 samples, respectively), two-group comparison tests were used. Based on these, 845 proteins were found to be differentially expressed, 356 with a fold-change (FC) of over 2, 183 under-expressed, and 173 overexpressed in LC. For example, several components of the basement membrane (e.g., collagens, nidogen-1, laminin subunit 3) were downregulated in tumor tissue, and proteins involved in calcium-ion binding (e.g., annexin-A3, S100A4) showed lower expression levels in tumor tissue than in tumor-adjacent tissue. Additionally, many proteins related to Ribonucleoprotein biogenesis and organization (small- and large ribosomal subunit proteins) were overexpressed in tumor tissue.

4.2.2. Lung Cancer Type-Specific Differences between Tumor and Adjacent Normal Tissue

Two-group comparisons were performed separately for all 4 LC types. This revealed that there are 78 proteins differentially expressed in all four LC types compared to adjacent tissue (*Figure 8*). The majority of alterations occurred in a group-specific manner: 61 proteins were differentially expressed only in AC, while 201, 119, and 44 proteins only in SCLC, SqCC, and LCC, respectively (*Figure 8*). For example, proteins connected to the hemostasis (e.g., fibrinogens), and proteins involved in the RHO protein signal transduction (e.g., apolipoproteins, CO1A2) showed lower expression levels only in SqCC tissue compared to tumor-adjacent tissue. Furthermore, several proteins related to splicing (splicing factors, and heterogeneous nuclear ribonucleoproteins) were significantly overexpressed only in SCLC tissue, while proteins related to microtubule organization (e.g., tubulin beta chains like TBB4B and TBB5, microtubule associated proteins like MAP4 and MAP1S) were also found to be highly overexpressed only in SCLC tissue.

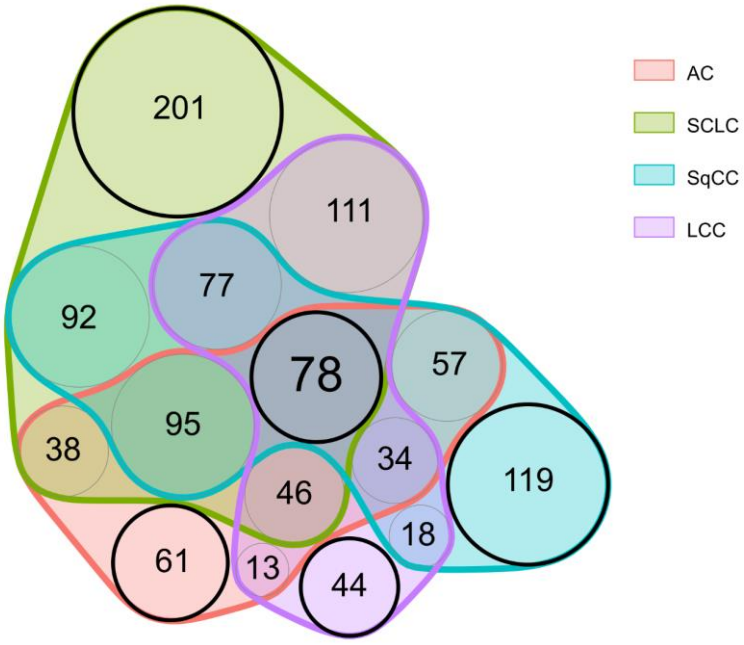

Figure 8: Venn diagram displaying the proteins differentially expressed in the different LC types compared to adjacent tissue. Figure made by the author with nVennR in R.

The specific groups mentioned before and highlighted in *Figure 8* are particularly noteworthy. First, the „central" 78 proteins might be useful in detecting LC, regardless of type. Second, the other „LC type-specific" proteins might be used for the detection of

different types of LC following further validation. Six examples are presented in *Figure 9* showing examples of proteins showing distinct behavior patterns.

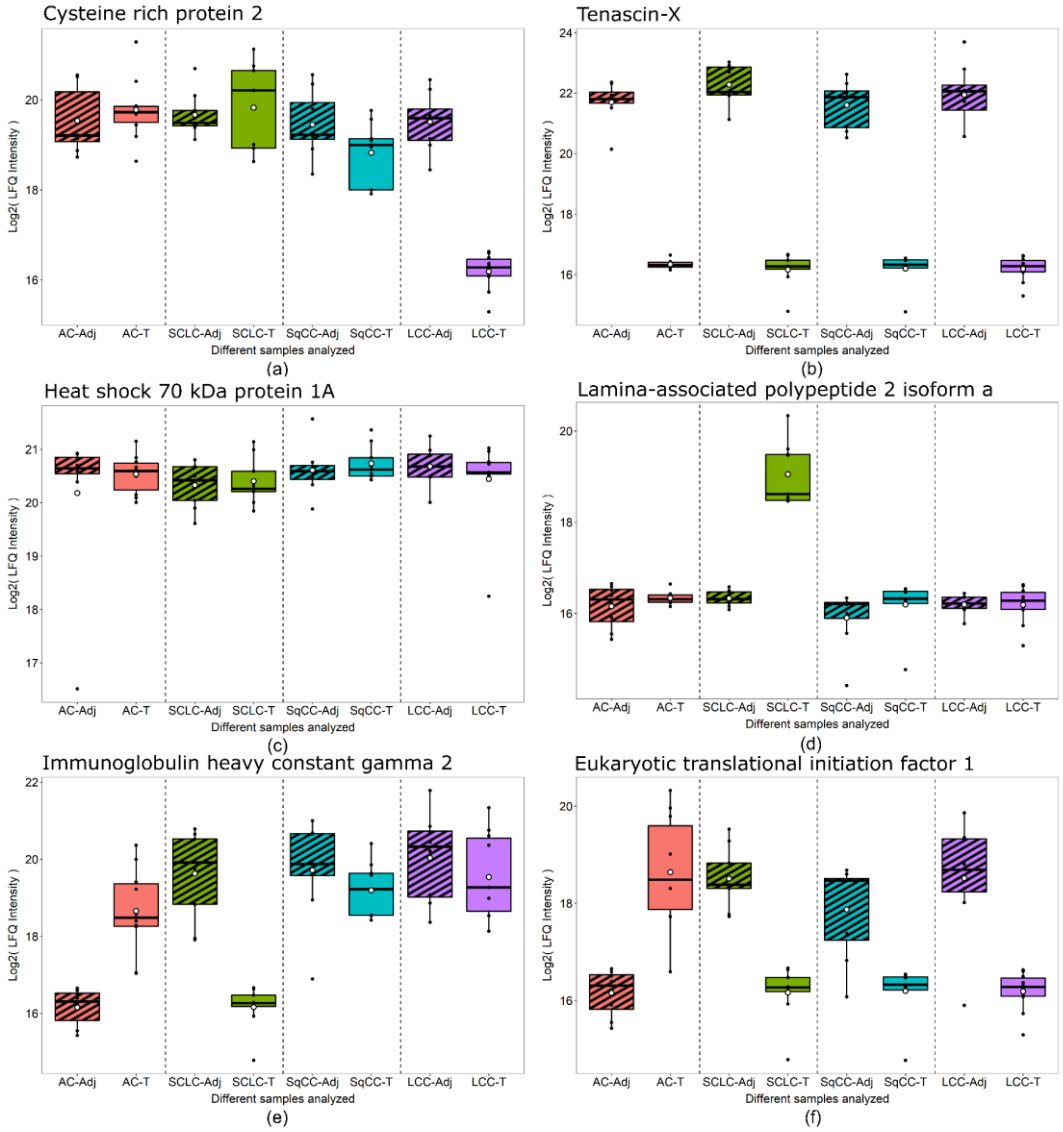

Figure 9: Box plots of protein examples for distinct expression patterns observed. (**a**) Cysteine-rich protein 2 showed under-expression in only LCC tumor tissue. (**b**) Tenascin-X was under-expressed in all tumor tissue types compared to adjacent tissue. (**c**) Heat shock 70 kDa protein 1A did not show differential expression between any tissue-pairs. (**d**) Lamina-associated polypeptide 2, isoform alpha was overexpressed in only SCLC tumor tissue. (**e**) Immunoglobulin heavy constant gamma 2 was overexpressed in AC tumor, while under-expressed in SCLC tumor compared to adjacent tissue. (**f**) Eukaryotic translational initiation factor 1 was overexpressed in AC tumor while under-expressed in all other tumor types compared to adjacent tissue. Boxes that represent adjacent tissue are striped, while tumor tissue are filled without pattern. Figure made by the author with ggplot2 in R.

Cysteine-rich protein 2 was under-expressed in only LCC tumor compared to the tumor-adjacent normal tissue regions and is subsequently part of the 44 proteins differentially expressed only in LCC tissue (*Figure 8*). Tenascin-X, on the other hand, was under-

expressed in all types of LC analyzed; thus, part of the 78 proteins differentially expressed in all tumor types (*Figure 8*). Heat shock 70 kDa protein 1A did not show differential expression but was highly abundant in all tissue types. Lamina-associated polypeptide 2, isoform alpha was overexpressed only in SCLC tumor tissue and is part of 201 proteins differentially expressed only in SCLC (*Figure 8*). Finally, IGHG2 and eukaryotic translational initiation factors were found to be differentially expressed in more than one LC types, however, not in a uniform direction.

4.2.3. Differences between the Different Types of Tumor Tissue

Proteins differentially expressed between the cancerous tissue regions of the four types of LC were identified using multiple sample and two-sample comparisons (for details see *Methods*), resulting in 571 proteins with altered expression. None of the proteins showed altered expression in all pairwise comparisons, but 23 of them showed differences in 5 out of 6 comparisons. There were several proteins with changes in expression in only one LC type, separating it from all the others (61 for AC, 99 for SCLC, 35 for SqCC, and 47 for LCC). For example, eukaryotic translation initiation factor 1 and matrix-metalloproteinase proteins (MMP2 and MMP19) were found to be upregulated in AC tissue compared to other LC types. IGHG2, and RAB10—a small GTPase related to Golgi vesicle transport—were significantly downregulated in SCLC tissue compared to other LC types. Fascin—an actin filament bundling protein—was found to be significantly overexpressed in SqCC tissue.

Hierarchical clustering based on the 571 differentially expressed proteins revealed that the molecular profiles of the different types of LC were markedly different, and the clustering of samples—except for 2—was in agreement with the pathological classification (*Figure 10*). Despite the histological classification, the three LCNEC samples clustered closely together with the 7 LCC samples, confirming that the phenotype is determined by a more complex molecular profile.

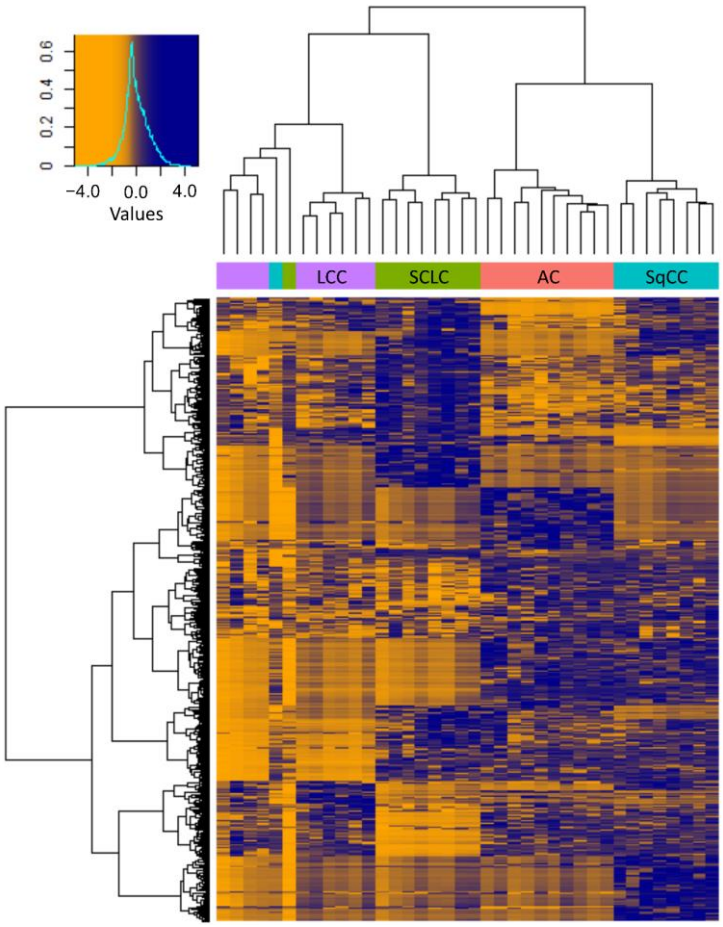

Figure 10: Heatmap with hierarchical clustering of the 571 proteins showing differential expression between the tumorous regions of the 4 LC types investigated (Protein LFQ Intensity values are Z-scored). The histogram in the top left corner shows the correspondence between color hues and values. Figure made by the author with gplots in R.

To identify the dysregulated biological processes in the tumorous regions compared to adjacent tissue, pre-ranked Gene Set Enrichment Analysis (GSEA) was applied. GSEA provides information on which biological terms are enriched based on a list of proteins similar to STRING (see *4.1.3. Biological Processes Altered in Prostate Cancer*); however, here, quantitative data for all proteins is considered. This was performed separately for all four LC types (based on effect sizes), then the results were compared to reveal gene sets enriched in LC in general, and in a type-specific manner. When discussing and comparing GSEA results, the Normalized Enrichment Score (NES) is reported occasionally, which is a normalized score that represents the degree to which a gene set is overrepresented on the top or the bottom of the gene list. The enriched gene sets could be separated into several major groups of biological processes: extracellular matrix (ECM) organization, assembly, regulation and adhesion; signaling cascades (e.g.,

Ca2+ dependent signaling, RHO signaling); processes involved in protein synthesis (transcription, translation, DNA, and RNA related processes); humoral and cell-mediated immune system processes; and transport processes (e.g., vesicular transport, endocytosis). The processes dysregulated in LC were similar in all types; however, there were several type-specific differences, especially between SCLC and NSCLC. This is not surprising based on *Figure 8*, which shows that SCLC has the largest number of unique dysregulated proteins out of all 4 types investigated. An excerpt of these pathways along with their NES values are summarized in *Table 3*.

Table 3: Excerpt of the dysregulated biological pathways along with Normalized Enrichment Scores (NES). *NA* values stand for *not applicable*.

Name	AC	SCLC	SqCC	LCC
KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	−2.03	−1.98	−3.05	−1.54
KEGG_FOCAL_ADHESION	−2.51	−2.36	−2.46	−1.72
GOCC_ACTIN_CYTOSKELETON	NA	−2.05	−1.44	−1.65
GOCC_COLLAGEN_CONTAINING_EXTRACELLULAR_MATRIX	−2.17	−2.5	−2.84	−1.7
GOMF_CALCIUM_ION_BINDING	NA	−2.31	NA	NA
GOBP_ALTERNATIVE_MRNA_SPLICING_VIA_SPLICEOSOME	NA	2.58	NA	NA
GOBP_BIOLOGICAL_ADHESION	NA	−2.39	−1.68	NA
GOBP_CELLULAR_MACROMOLECULE_BIOSYNTHETIC_PROCESS	1.69	2.96	1.95	1.86
GOBP_CYTOPLASMIC_TRANSLATION	1.73	2.4	1.87	1.91
GOBP_ENDOCYTOSIS	NA	−2.39	−2.71	−1.65
GOBP_G_PROTEIN_COUPLED_RECEPTOR_SIGNALING_PATHWAY	NA	−1.81	−2.22	NA
GOBP_GENE_SILENCING_BY_RNA	1.64	2.45	1.72	1.35
GOBP_GLYCEROLIPID_METABOLIC_PROCESS	−2.15	−1.69	−1.85	NA
GOBP_HEMOSTASIS	−1.83	−1.86	−2.18	NA
GOBP_HUMORAL_IMMUNE_RESPONSE	NA	−1.85	−2.44	NA
GOBP_LYMPHOCYTE_MEDIATED_IMMUNITY	NA	−1.99	−2.17	NA
GOBP_MRNA_METABOLIC_PROCESS	1.61	3.49	1.78	1.33
GOBP_NCRNA_METABOLIC_PROCESS	1.63	2.42	1.9	2.11
GOBP_PEPTIDE_BIOSYNTHETIC_PROCESS	1.81	2.83	1.89	1.84
GOBP_PYRUVATE_METABOLIC_PROCESS	1.75	NA	1.55	1.43
GOBP_REGULATION_OF_CHROMOSOME_ORGANIZATION	NA	2.74	1.58	1.66
GOBP_REGULATION_OF_VESICLE_MEDIATED_TRANSPORT	NA	−1.97	−2.13	NA
GOBP_RHO_PROTEIN_SIGNAL_TRANSDUCTION	NA	−1.45	−2.13	NA
GOBP_RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS	1.79	2.88	2.05	1.94

5. Discussion

5.1. Proteomics and *N*-Glycoproteomics of Different Grades of Prostate Cancer

As the focus of this research project was on finding potential biomarkers through exploring alterations in the glycosylation between healthy and PCa tissues combined with proteomics data, only glycoproteins displaying significant changes are discussed individually. For these, the differences in protein expression and glycosylation are both reported, and they are compared to relevant previous studies on PCa or cancer in general. Furthermore, the most significant biological processes are also discussed.

The PPI network analysis provides information about biological processes altered in PCa. The underexpressed proteins were mostly associated with cellular component organization (34 out of 51) and various processes connected to adhesion e.g., the KEGG term "*Focal adhesion*" and the GO term "*cell adhesion*", and muscle contraction e.g., the KEGG term "*Vascular smooth muscle contraction*" and the GO term "*muscle contraction*". Focal adhesion has been confirmed to be heavily involved in cancer progression, [146] while smooth muscle cells have been reported to be involved in PCa and BPH. [147] The overexpressed proteins, on the other hand, were primarily associated with metabolic processes (60 out of 72) with the GO terms "*localization*" and "*regulation of gene expression*" involving the most proteins. While altered localization of macromolecules in a cell (e.g., proteins [148]) can reportedly drive tumor development and metastasis, aberrant gene expression is known to be the principal cause of cancer. [149]

All glycoproteins with significant glycosylation changes were quantified in the proteomics part of the study by MaxQuant, but not all of them showed differential expression between normal and PCa tissues. This suggests that altered glycosylation does not necessarily indicate glycoprotein-wise differential expression. Furthermore, neither of the metrics (sialylation, fucosylation, galactosylation) used for the characterization of glycosylation showed significant overall changes between PCa and normal tissue. Regarding cellular localization, all the glycoproteins with significant glycosylation changes were primarily of extracellular origin, most of them were associated with the ECM and consequently, the tumor microenvironment (TME), which is known to heavily influence cancer initiation, progression, and invasion. [150]

There are several changes in glycosylation that are known to widely occur in cancer. These include increased and altered sialylation, increased glycan branching (number of antennae on a glycan) and altered fucosylation. [26,151] Furthermore, there have been many PCa glycome-specific changes reported before, [152] e.g., the expression of oligomannosidic glycans in the tumor region in late-stage PCa. [153] These changes, however, reflect only overall tendencies, they are not necessarily true for all of the glycosylation sites, as the results clearly demonstrate.

In previous studies, serum sialylation has been found the be correlated with pathological grade and elevated sialic acid levels with PCa bone metastasis. [154] In tissues, however, overall sialylation levels have been reported to be constant across different grades of cancer. [155] The results suggest the same, the average sialylation levels were similar throughout the different sample groups, but there were significant individual differences detected on several glycosylation sites. Most of them showed a decrease in sialylation except for CO6A2 N785, which showed an overall increase and significant differences between the different pathological grade groups. Also, proteomics results revealed that CO6A1, CO6A2, and CO6A3 expression levels significantly changed with PCa progression in a similar manner. This is highlighted by the fact that CO6A1 has been reported to have an important role in tumor growth, and the molecular etiology of castration-resistant prostate cancer (CRPC). [156]

Apart from serum, PCa cell lines have also been used before to identify diagnostic markers, and site-specific changes in fucosylation have been reported in PC3 and LNCap cell lines. [157] This aligns with the findings, as it was also found that fucosylation increased in PCa on multiple glycosites. Also, PPAP has been demonstrated to have a significant effect on PCa cell growth, [158] and it has been hypothesized to have higher site-specific fucosylation levels in PCa patients. [157] This is supported by the data: the average fucosylation level of PPAP N94 increased from 47 to 83% in PCa.

POSTN has been reported to be upregulated in aggressive PCa, [159] but significant changes in glycosylation have not been reported yet. The proteomics results reaffirmed, that POSTN is overexpressed in PCa, and significant changes were also detected in both fucosylation and sialylation on POSTN N599, an increase from 24 to 72% and a decrease from 83 to 44% respectively, highlighting their possible importance.

Prostate tissue is known to be a rich reservoir of prothrombin, [160] the precursor of thrombin, which has been reported to promote prostate tumor growth, increase tumor cell seeding, and stimulate angiogenesis. [161,162] It was found that the sialylation of THRB N121 was downregulated significantly in PCa, moreover, with the largest relative difference.

Alterations of serum IgG glycosylation has been reported in many diseases, including PCa, [163] and IgG1 has been suggested as a potential target for PCa treatment. [164] It was found that both IGG1 N299 and IGHG2 N176 show decreased overall galactosylation by 6.3% and 8.7% respectively. This is in line with previous studies, where one of the major differences reported was the decrease of terminal galactosylation in PCa compared to either healthy or benign prostatic disease patients. [165] The data also shows reduced sialylation on both IGG1 N299 and IGHG2 N176 by 2.1% (corresponding to a relative change of 21.3% and 26.8% respectively), which is also in agreement with literature as reduced sialylation has been described as a major alteration in PCa compared to healthy individuals. [166]

Another glycoprotein with significant site-specific glycosylation changes was MFAP4, which has been reported to be involved in several cancers and may function as a tumor suppressor in PCa. [167] MFAP4 has been documented to have altered glycosylation in pancreatic adenocarcinoma, [168] however, not in PCa. The results revealed that both sites of MFAP4 showed modified glycosylation in PCa: decreased sialylation on N87 and increased expression of the glycan N4H5S1F1 on N137. The latter glycoform might be a useful indicator in detecting PCa at an early stage, as this increased expression was detected between normal and G1 samples.

Most of the glycoproteins discussed above can be found in the Human Protein Atlas [169] (apart from IGG1 and IGHG2) and are categorized in the Pathology Atlas based on *Prognostic summary* and *Cancer specificity*. Apart from PPAP, which is a protein specific to PCa, all of them are unfavorable prognostic markers in certain types of cancer (in most cases renal cancer) which suggests that these glycoproteins are heavily involved in cancer progression.

It is also important to note, that these glycoproteins have been detected in biofluids previously. All glycoproteins discussed above with the exception of POSTN have been detected in urine, [170] while POSTN has been detected in serum samples [171] of PCa

patients. This suggests their potential usefulness as a clinical marker. Whether the alterations in the glycosylation of these proteins are PCa specific or not, needs further investigation, especially in the context of their biomarker status.

5.2. Proteomics of the Four Lung Cancer Types

The four most common types of lung cancer (AC, SCLC, SqCC, and LCC) have been investigated through MS-based proteomic analysis of FFPE tissue sections following on-surface digestion. Based on a thorough literature search, this was the first pilot study comparing these four types of LC in the same cohort. On-surface digestion has the advantage that small tissue areas corresponding to cancerous, and cancer adjacent regions can be investigated from the same FFPE material. However, there is a compromise, as analyzing small tissue areas results in fewer proteins identified compared to bulk tissue analysis. Previous experiences indicate that this approach can be successfully used for FFPE tissues and biopsies. [133] The results were compared to a study specifically focused on lung AC, and several of the 78 “central” proteins altered in all LC types were detected in both studies and the direction of changes were in agreement. [172] Furthermore, two of them (annexin A3 and tenascin-X) were previously reported with altered blood levels between LC patients and healthy controls. [173]

The grouping of samples was based on histological appearance. Three LCC samples were LCNEC based on classification, as they expressed at least one out of several neuroendocrine markers. The phenotype of LCNEC cells, however, is influenced by their entire molecular profile, not just the expression of specific markers, therefore they were grouped with LCC samples. Furthermore, based on hierarchical clustering (*Figure 10*), these three LCNEC samples clustered with the LCC samples, confirming the validity of the grouping based on large cell morphology.

Proteins with altered expression levels—comparing any of the types or adjacent samples—will be discussed in the context of the results of the gene set enrichment analysis which was performed to identify altered biological processes. The biological pathways discussed and the direction of their dysregulation are summarized in *Table 3*. Considering that the cancerous and tumor-adjacent normal regions of these tissues were analyzed, some proteins and biological processes which are otherwise dysregulated in cancerous tissue compared to healthy controls may not have been revealed. The main biological processes identified as disrupted—and discussed further in detail—in any of

the types of LC analyzed include extracellular matrix remodeling, altered adhesion, signaling cascades, immune response, coagulation, protein biosynthetic processes, metabolic processes, and vesicular transport.

ECM remodeling processes are prevalent in cancer, creating a microenvironment that promotes tumorigenesis and metastasis. [174] It has been identified through GSEA analysis that the gene ontology cellular compartment (GOCC) gene set *"Collagen containing ECM"* was significantly and heavily suppressed in all four LC types investigated (with an average NES value of over 2). This—in the most part—is due to the altered expression of major ECM constituent protein families: collagens, laminins, nidogens, proteoglycans, and matrix metalloproteinases.

Altogether 17 collagen proteins were identified. None of them showed altered expression between the four cancer tissue types, but most of them were downregulated in cancerous tissue in general, compared to adjacent tissue. Several types of collagens play a complex role in tumor proliferation, invasion, angiogenesis, and metastasis. [175,176] Elevated levels of CO1A1 have been reported to be associated with chemoresistance and poor progression-free survival in metastatic lung cancer. [177] Nidogen-1 (NID1) was downregulated in all tumor types, while NID2 was downregulated only in SCLC, compared to adjacent tissue. It has been previously reported that in serum samples, the degradation of NID1 is associated with NSCLC. It was also reported that NID1 enhances cell proliferation, migration, invasion, and promotes lung metastasis of breast cancer and melanoma. [178]

Seven subunits of laminin have been identified. Subunit gamma-1 (LAMC1) was downregulated in all four cancer types, while all other subunits showed lower expression levels in SCLC; LAMC2, LAMA5, LAMB2 in LCC; and LAMA5, LAMB2, LAMA3 in AC as well. Additionally, LAMB1 was under-expressed in SCLC and LCC but overexpressed in SqCC, and it also showed significant differences between SqCC and all other tumor tissues. It has been suggested that LAMC2 promotes metastasis in AC [179], while another study demonstrated the significance of circulating LAMC2 as a prognostic marker in SCLC, especially for early-stage cancer. [180]

Out of the nine proteoglycans identified, versican was overexpressed in AC, while perlecan, decorin, prolargin, and mimecan were all under-expressed in SCLC and LCC. Agrin and lumican were downregulated in SCLC: biglycan showed lower expression

levels in AC, SCLC, and SqCC. The versatility of changes in proteoglycan expression levels in the different types of LC is in line with previous studies, which report that their role in cancer is highly context dependent. [181,182]

Five different matrix metalloproteinases (MMPs) have been identified. MMP-2 was upregulated in AC compared to both other cancer types and adjacent tissue, making it a promising target for future studies. MMP-12 was significantly overexpressed in SCLC and LCC, MMP-19 was overexpressed in AC but under-expressed in SCLC. Based on previous studies, the overexpression of MMP2 is associated with tumor differentiation, invasion, angiogenesis, and metastasis. [183]

GSEA analysis revealed that many gene sets involved in adhesion and cytoskeleton remodeling were dysregulated. [184–186] The gene ontology biological process (GOBP) gene set *"Regulation of cell adhesion"* (and related terms such as cell-matrix, and cell-substrate adhesion) was suppressed in SCLC and SqCC; the gene ontology molecular function (GOMF) gene sets *"Actin binding"* in SCLC, SqCC and LCC; and *"Collagen, and Integrin binding"* in AC, SCLC, and SqCC. The GOCC gene sets involved in cytoskeleton formation were also suppressed, e.g., *"Actin cytoskeleton"* in SCLC, SqCC, and LCC, and *"Microtubule cytoskeleton"* in SCLC. Finally, the KEGG gene set *"Focal adhesion"*—a well-known process involved in cancer metastasis [187]—was found to be heavily suppressed in all types of LC. These changes are due to the altered expression (mostly under-expression) of proteins involved in cytoskeleton formation and cell adhesion, e.g., actins, tubulins, microtubule-associated proteins, spectrins, and galectins. Actin, cytoplasmic 2 (ACTG) and actin, alpha cardiac muscle 1 (ACTC) were found to be significantly under-expressed in SCLC and LCC. Furthermore, most of the identified actin-related proteins, such as myosins and filamins, were heavily downregulated in SCLC tissue. On the other hand, fascin (FSCN1)—an actin filament bundling protein—was significantly overexpressed in SqCC tissue compared to the other cancer types. The overexpression of FSCN1 in NSCLC has been previously reported to be associated with tumor growth, migration, invasion, and metastasis. [188,189]

All of the five identified tubulins showed altered expression in the different types of LC compared to adjacent tissue. Changes in the expression of microtubule-associated proteins (MAPs) were also identified. Microtubule-associated protein 1B (MAP1B) was significantly upregulated in SCLC tissue compared to the different types of NSCLC and

adjacent tissue. In addition to MAP1B, other dysregulated microtubule-associated proteins were also found, MAP1S and MAP4. MAPs have been previously reported to be commonly overexpressed in cancer. [184,185]

Spectrins were also found to be altered in LC, although there was a big difference between SCLC and NSCLC in this regard. All of them showed under-expression in SCLC while only 1 in each of the other NSCLC types. Furthermore, spectrin beta chain (SPTB) also showed differential expression between SCLC and all other cancer tissue. Ankyrin-1 (ANK1)—a protein that regulates cell shape and membrane integrity together with spectrins—was also found to be downregulated in all LC types. It has been previously reported that ANK1 and spectrins—such as SPTB and SPTA1—were under-expressed in AC compared to paired non-malignant tissue. [190]

Two galectins (LEGs) were identified, and it has been found that LEG1 was significantly under-expressed in SCLC compared to SCLC adjacent tissue. Expression levels of LEG3 in SCLC and LCC tumor were significantly lower than in AC, SqCC, and adjacent tissue. LEG1 promotes tumor cells invasion and migration; it is also a potential prognostic marker in early-stage NSCLC [191] and is feasible for the promotion of chemoresistance in AC. [192] Furthermore, LEG1 and LEG3 are associated with LC and correlated with tumor invasion, migration, metastasis, and progression. [192]

Many signaling pathways have been identified as dysregulated in LC tissue. The *"Mammalian target of Rapamycin 1 (MTORC1) signaling pathway"*—represented by the hallmark gene set—was found to be elevated in all four types of LC. The MTORC1 pathway plays an important role in regulating fatty acid (FA) metabolism, and, subsequently, the production of ATP. Furthermore, it was found that ATP-citrate synthase (ACLY)—a key enzyme involved in acetyl-CoA synthesis regulated by MTORC1—is significantly overexpressed in AC. The overexpression of ACLY has been observed in multiple tumor types before. [193] It has also been shown that targeting ACLY significantly reduces the growth of lung and prostate tumor xenografts. [194]

The GOBP gene set *"RHO protein signal transduction"* was found to be significantly suppressed in SqCC. From the gene set, three corresponding proteins were found to be highly under-expressed in SqCC compared to adjacent tissue: apolipoprotein A-I (APOA1), apolipoprotein E (APOE), and collagen alpha-2(I) chain (CO1A2). APOA1 has been previously reported to be under-expressed in NSCLC serum and associated with

metastasis and poor prognosis. [195] APOE is an important marker of cancer, its overexpression in NSCLC has been reported before. Downregulation of CO1A2 has also been reported previously in NSCLC. [196]

The GOMF gene set *"Calcium ion binding"* was found to be significantly and highly suppressed in SCLC. This is, in a large part, due to the changes detected in the expression levels of annexins—a large family of Ca2+-binding proteins involved in signaling processes. Seven different Annexins were identified with highly diverse expression profiles between the different types of LC. Annexin A3 (ANXA3) was found to be under-expressed compared to adjacent tissue in all types of LC. Aberrant expression of ANXA3 has been previously reported to promote tumor cell proliferation, invasion, metastasis, angiogenesis, and therapy resistance. [197] ANXA4 was under-expressed in all types except for AC. Previous studies suggest that ANXA4 enhances tumor invasion and promotes anti-tumor drug resistance. [198] The data also shows that ANXA2 was under-expressed in SCLC and LCC. Out of the seven annexins quantified, five showed differential expression in SCC, three in LCC, two in SqCC, and only one in AC. Protein S100-A4—another Ca2+-binding protein—was found to be under-expressed in all types of tumor tissue. It was previously reported to be downregulated in tumor cells compared to stromal non-tumor cells, confirming the results. [199]

Immune system processes are often deregulated in cancer. It has also been demonstrated that inflammatory immune cells are essential players in cancer-related inflammation. [200] Multiple processes involved in both humoral and cell-mediated immune responses were identified as altered in the different LC types analyzed. The KEGG gene set *"Complement and coagulation cascade"* was found to be suppressed in all four types of LC tumor tissue. On the other hand, humoral immune response and biological processes under cell-mediated immune response—*"Leukocyte mediated immunity"*, *"Lymphocyte mediated immunity"*, *"B cell mediated immunity"*, and *"Phagocytosis"*—were suppressed only in SCLC and SqCC. These LC type-specific differences were highlighted by the differences in the expression of the 12 lgG proteins identified. Nine of them were differentially expressed in SCLC, while only three in AC, four in SqCC, and two in LCC. Furthermore, there were substantial differences in the direction of these changes. Previous studies have also reported differences in IgG expression and subclass distribution in LC patients. [201,202]

Pathophysiological changes in the body can affect the hemostatic system. Specifically, LC has been reported as one of the cancers with the most risk of developing venous thromboembolism. [203] The GOBP gene sets, *"Hemostasis"* and *"Regulation of coagulation"* were suppressed in AC, SCLC, and SqCC, however, in SqCC, with an especially high NES (2.2 and 2.8, respectively). This is primarily due to the differences in fibrinogen expression levels. Three fibrinogens were identified: fibrinogen alpha chain (FIBA), fibrinogen beta chain (FIBB), and fibrinogen gamma chain (FIBG). In SqCC, all of them were under-expressed, while none of them in the other types of LC. The importance of fibrinogens in cancer is highlighted by the fact that plasma fibrinogen levels have been previously suggested as prognostic markers in NSCLC, combined with the neutrophil-to-lymphocyte ratio. [204,205]

Increased protein synthesis and subsequent rapid cell growth is one of the hallmarks of cancer, which can be activated through the dysregulation of various biological processes. [206] Several of these were identified connected to four major umbrella terms: translation, mRNA, ribonucleoprotein, and chromosome-related processes.

The GOBP gene set *"Cytoplasmic translation"* was found to be activated in all four types; on the other hand, *"Translational initiation"* was elevated only in AC, SCLC, and SqCC. This is explained by the differences in the expression levels of the 16 initiation factors (IFs) identified. The number of IFs differentially expressed was the lowest in LCC compared to other types of LC, with significantly lower effect sizes. Three initiation factors showed differential expression in all four types: eukaryotic initiation factor 4A-I (IF4A1) and eukaryotic translation initiation factor 6 (IF6) showed overexpression in all four; however, eukaryotic translation initiation factor 1 (EIF1) was under-expressed in SCC, SqCC and LCC, but overexpressed in AC. EIF1 also showed a difference in expression between AC and all other types of LC. Dysregulation, specifically the overexpression of the initiation factors has been reported previously to occur frequently in cancer. [207]

Multiple GOBP gene sets involved in mRNA processing have been identified—elevated in all four types: *"mRNA and ncRNA processes"*, *"mRNA and ncRNA metabolic processes"*, and *"Gene silencing by RNA"*. However, gene sets related to splicing: *"Alternative mRNA splicing"*, and *"Regulation of mRNA splicing via spliceosome"*, were elevated only in SCLC. This can be explained by looking at the expression levels of

splicing factors (SFs), and heterogeneous nuclear ribonucleoproteins (hnRNPs). Out of the 16 SFs identified overexpression was detected in 13 in SCLC, while for NSCLC subtypes, this number was much lower. Regarding the 15 hnRNPs, 13 showed significant overexpression in SCLC, while—similarly to SFs—this number was much lower in the NSCLC subtypes. mRNA-related processes are potential therapeutic targets for cancer, as it has been reported that alterations in mRNA can contribute to the initiation and progression of cancer. [208] The activation of GOBP gene sets *"Ribonucleoprotein biogenesis"* and *"Ribonucleoprotein subunit organization"* was detected in all four types of LC. 18 small ribosomal subunits (40S) and 27 large ribosomal subunits (60S) were identified, with many of them overexpressed in all four types of LC. The altered expression of Ribosomal proteins has been linked to multiple LC-related processes, e.g., the downregulation of 60S ribosomal protein L3 (uL3) to drug resistance in LC cells, and 40S ribosomal protein S6 (eS6) was found to be overexpressed in NSCLC. [209,210]

Other important processes connected to protein biosynthesis found was the activation of *"Regulation of chromosome organization"* (in SCLC, SqCC, and LCC), *"Chromatin organization"* (in SCLC), *"Positive regulation of DNA metabolic process"*, and *"Positive regulation of DNA repair"* (in SCLC, SqCC, and LCC). It has been previously reported that chromosome loop structures can be altered in cancer and contribute to oncogene dysregulation. [211] Furthermore, it has been suggested that DNA repair pathways could be used as therapeutic targets in both SCLC and NSCLC. [212]

Metabolic remodeling is necessary for all stages of tumor development. [206] Several dysregulated GOBP gene sets involved in metabolic processes have been identified by GSEA. *"Glycerolipid metabolic process"* was elevated in AC, SCLC, and SqCC; *"Pyruvate metabolic process"* in AC, SqCC, and LCC; and *"Cellular carbohydrate metabolic process"* in AC and SqCC. This suggests that further investigation of the differences in metabolism between the different types of LC is reasonable, which is further highlighted by previous studies reporting that abnormal glycolysis and lipid metabolism have a significant role in the development of LC. [213]

Vesicle related GOCC and GOBP gene sets (e.g., *"Endocytic vesicle"*, *"Regulation of vesicle mediated transport"*, *"Endocytosis"*) were suppressed in SCLC and SqCC. It has been previously reported that LC-derived extracellular vesicles mediate epithelial-mesenchymal transition by the transfer of vimentin—which was also found to be

differentially expressed in SCLC and SqCC—and regulate angiogenesis, activate cancer-associated fibroblasts, and mediate metastasis. [214,215] *"Golgi vesicle transport"* was also elevated in SqCC tissue but suppressed in SCLC. RAB10—a small GTPase of the Ras superfamily—is a key component of this process, which was found to be upregulated in SqCC and downregulated in SCLC. Rab GTPase proteins have been previously reported to have diverse roles in cancer progression as both oncoproteins and tumor suppressors. [216]

6. Conclusions

6.1. Proteomics and *N*-Glycoproteomics of Different Grades of Prostate Cancer

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

6.2. Proteomics of the Four Lung Cancer Types

In conclusion, the proteomic analysis of FFPE tissue from four different types of lung cancer resulted in the identification of several disrupted biological processes in all investigated cancer types, such as the degradation of the basement membrane and suppression of the complement and coagulation cascade, as well as the activation of the MTORC1 signaling pathway. Furthermore, the differences between SCLC and NSCLC samples were larger than between the three distinct NSCLC subtypes. Dysregulated pathways differentiating SCLC from NSCLCs include suppressed regulation of cell adhesion, actin filament-based processes, and calcium ion binding. Overexpression of splicing factors and heterogeneous nuclear ribonucleoproteins suggest that biological processes related to splicing are more affected in SCLC. Additionally, the expression of several proteins showed changes only in one LC type, such as the overexpression of FSCN1 in SqCC or the downregulation of NID2 in SCLC. These results correlated well with previous studies analyzing individual NSCLC types and tumor adjacent tissue, even though the tissue areas available were smaller. The specific molecular signatures connected to LC in general and to specific LC types described in detail might be attractive targets for further in-depth investigations may have potential diagnostic and prognostic value.

DOI:10.14753/SE.2023.2877

7. Summary

Two projects are summarized in this thesis: the proteomic and *N*-glycoproteomic analysis of cancerous and healthy prostate tissue microarray samples; and the proteomic analysis of different types of lung cancer tissue. Although in several ways similar, their focus was different—in line with current cancer-specific clinical demands and challenges.

The primary aim of the proteomic and *N*-glycoproteomic analysis of prostate TMAs was the integration and application of previous research for the identification of molecular alterations between PCa and Normal tissue with potential diagnostic value. The combination of results from proteomics and *N*-glycoproteomics indicate that alterations in glycosylation occur primarily at the glycosite level—while overall glycosylation may remain unaffected—often without a change in protein expression. Several glycoproteins were identified with altered glycosylation; the changes in fucosylation on CO6A2 N785 were especially noteworthy with regards to potential early cancer detection. This suggests that the investigation of these potential prognostic markers, and the identification of their exact roles is reasonable and could lead to further advancement in understanding the function of *N*-glycosylation in cancer development and PCa diagnosis.

The main objective of the proteomic analysis of different types of lung cancer tissue was to identify dysregulated biological processes by analyzing and comparing the protein expression of the four main types of LC (small cell lung cancer, adenocarcinoma, squamous cell carcinoma, and large cell carcinoma) with their adjacent normal regions and each other. Several such processes were identified, the majority of them dysregulated in all four LC types. There were, however, LC type-specific differences as well; differences were larger between SCLC and NSCLC than between the three distinct NSCLC subtypes. Dysregulated pathways differentiating SCLC from NSCLCs include the suppressed regulation of cell adhesion, actin filament-based processes, and calcium ion binding. Additionally, the expression of several proteins showed changes in only one LC type, such as the overexpression of FSCN1 in SqCC, or the downregulation of NID2 in SCLC. These results correlated well with previous studies analyzing individual NSCLC types and tumor adjacent tissue. The specific molecular signatures connected to LC in general and to specific LC types described in detail might be attractive targets for further in-depth investigations and may have potential diagnostic and prognostic value.

8. References

1. Global health estimates: Leading causes of death, 2000–2019, by Cause, by Age, by Country and by Region, 2000-2019. World Health Organization 2020.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–249.
3. Damjanov I, Fan F. Cancer grading manual. Springer; 2007.
4. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, Postmus PE, Rusch V, Sobin L, Committee IA for the S of LCIS. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. Journal of thoracic oncology. 2007;2(8):706–714.
5. Burke HB. Outcome prediction and the future of the TNM staging system. Vol. 96, Journal of the National Cancer Institute. Oxford University Press; 2004. p. 1408–9.
6. Ashrafizadeh M, Aghamiri S, Tan SC, Zarrabi A, Sharifi E, Rabiee N, Kadumudi FB, Pirouz AD, Delfi M, Byrappa K. Nanotechnological approaches in prostate cancer therapy: Integration of engineering and biology. Nano Today. 2022;45:101532.
7. Humphrey PA. Histopathology of prostate cancer. Cold Spring Harb Perspect Med. 2017;7(10):a030411.
8. Egevad L, Delahunt B, Srigley JR, Samaratunga H. International Society of Urological Pathology (ISUP) grading of prostate cancer–An ISUP consensus on contemporary grading. Vol. 124, Apmis. Wiley Online Library; 2016. p. 433–435.
9. Cooperberg MR, Broering JM, Carroll PR. Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. JNCI: Journal of the National Cancer Institute. 2009;101(12):878–887.
10. Hernandez DJ, Nielsen ME, Han M, Partin AW. Contemporary evaluation of the D'amico risk classification of prostate cancer. Urology. 2007;70(5):931–935.
11. Merriel SWD, Pocock L, Gilbert E, Creavin S, Walter FM, Spencer A, Hamilton W. Systematic review and meta-analysis of the diagnostic accuracy of prostate-

specific antigen (PSA) for the detection of prostate cancer in symptomatic patients. BMC Med. 2022;20(1):1–11.

12. Bergstralh EJ, Roberts RO, Farmer SA, Slezak JM, Lieber MM, Jacobsen SJ. Population-based case-control study of PSA and DRE screening on prostate cancer mortality. Urology. 2007;70(5):936–941.
13. Stabile A, Giganti F, Rosenkrantz AB, Taneja SS, Villeirs G, Gill IS, Allen C, Emberton M, Moore CM, Kasivisvanathan V. Multiparametric MRI for prostate cancer diagnosis: current status and future directions. Nat Rev Urol. 2020;17(1):41–61.
14. Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala MH, Briganti A, Budäus L, Hellawell G, Hindley RG. MRI-targeted or standard biopsy for prostate-cancer diagnosis. New England Journal of Medicine. 2018;378(19):1767–1777.
15. Drost FH, Osses DF, Nieboer D, Steyerberg EW, Bangma CH, Roobol MJ, Schoots IG. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. Cochrane Database of Systematic Reviews. 2019;(4).
16. Mottet N, Bellmunt J, Briers E, van den Bergh RCN, Bolla M, van Casteren NJ, Cornford P, Culine S, Joniau S, Lam T. Guidelines on prostate cancer. European Association of Urology. 2015;56:e137.
17. Popper HH. Progression and metastasis of lung cancer. Cancer and Metastasis Reviews. 2016;35(1):75–91.
18. Feldman R, Kim ES. Prognostic and predictive biomarkers post curative intent therapy. Ann Transl Med. 2017;5(18).
19. Bogos K, Kiss Z, Gálffy G, Tamási L, Ostoros G, Sárosi V, Vokó Z, Nagy B. A tüdőrák hazai epidemiológiai adatai új megközelítésben. Magy Onkol. 2020;64(3):175–181.
20. Inamura K. Lung cancer: understanding its molecular pathology and the 2015 WHO classification. Front Oncol. 2017;7:193.
21. Waqar SN, Morgensztern D. Treatment advances in small cell lung cancer (SCLC). Pharmacol Ther. 2017;180:16–23.

22. Oronsky B, Reid TR, Oronsky A, Carter CA. What's new in SCLC? A review. Neoplasia. 2017;19(10):842–847.
23. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422(6928):198–207.
24. Yates III JR. Mass spectrometry: from genomics to proteomics. Trends in genetics. 2000;16(1):5–8.
25. Aebersold R, Goodlett DR. Mass spectrometry in proteomics. Chem Rev. 2001;101(2):269–296.
26. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH. Essentials of Glycobiology [internet]. 2015;
27. Esmail S, Manolson MF. Advances in understanding N-glycosylation structure, function, and regulation in health and disease. Eur J Cell Biol. 2021;100(7–8):151186.
28. Petrović T, Lauc G, Trbojević-Akmačić I. The Importance of Glycosylation in COVID-19 Infection. In: The Role of Glycosylation in Health and Disease. Springer; 2021. p. 239–264.
29. del Val IJ, Kontoravdi C, Nagy JM. Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns. Biotechnol Prog. 2010;26(6):1505–1527.
30. Liu L. Antibody glycosylation and its impact on the pharmacokinetics and pharmacodynamics of monoclonal antibodies and Fc-fusion proteins. J Pharm Sci. 2015;104(6):1866–1884.
31. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. Nat Genet. 2003;33(3):311–323.
32. Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. Proteomics: technologies and their applications. J Chromatogr Sci. 2017;55(2):182–196.
33. Hu A, Noble WS, Wolf-Yadlin A. Technical advances in proteomics: new developments in data-independent acquisition. F1000Res. 2016;5.
34. Gillet LC, Leitner A, Aebersold R. Mass spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing. Annual review of analytical chemistry. 2016;9:449–472.

35. Kelleher NL. Peer reviewed: top-down proteomics. Anal Chem. 2004;76(11):196-A.
36. Catherman AD, Skinner OS, Kelleher NL. Top down proteomics: facts and perspectives. Biochem Biophys Res Commun. 2014;445(4):683–693.
37. Toby TK, Fornelli L, Kelleher NL. Progress in top-down proteomics and the analysis of proteoforms. Annu Rev Anal Chem (Palo Alto Calif). 2016;9(1):499.
38. Melby JA, Roberts DS, Larson EJ, Brown KA, Bayne EF, Jin S, Ge Y. Novel strategies to address the challenges in top-down proteomics. J Am Soc Mass Spectrom. 2021;32(6):1278–1294.
39. Royle L, Radcliffe CM, Dwek RA, Rudd PM. Detailed structural analysis of N-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions. Glycobiology protocols. 2007;125–143.
40. Jensen PH, Karlsson NG, Kolarich D, Packer NH. Structural analysis of N-and O-glycans released from glycoproteins. Nat Protoc. 2012;7(7):1299–1310.
41. Segu Z, Stone T, Berdugo C, Roberts A, Doud E, Li Y. A rapid method for relative quantification of N-glycans from a therapeutic monoclonal antibody during trastuzumab biosimilar development. In: MAbs. Taylor & Francis; 2020. p. 1750794.
42. Nwosu CC, Strum JS, An HJ, Lebrilla CB. Enhanced detection and identification of glycopeptides in negative ion mode mass spectrometry. Anal Chem. 2010;82(23):9654–9662.
43. Cao L, Qu Y, Zhang Z, Wang Z, Prytkova I, Wu S. Intact glycopeptide characterization using mass spectrometry. Expert Rev Proteomics. 2016;13(5):513–522.
44. Xue Y, Xie J, Fang P, Yao J, Yan G, Shen H, Yang P. Study on behaviors and performances of universal N-glycopeptide enrichment methods. Analyst. 2018;143(8):1870–1880.
45. Kim BJ, Dallas DC. Systematic examination of protein extraction, proteolytic glycopeptide enrichment and MS/MS fragmentation techniques for site-specific profiling of human milk N-glycoproteins. Talanta. 2021;224:121811.

46. Bapiro TE, Richards FM, Jodrell DI. Understanding the complexity of porous graphitic carbon (PGC) chromatography: modulation of mobile-stationary phase interactions overcomes loss of retention and reduces variability. Anal Chem. 2016;88(12):6190–6194.
47. Goumenou A, Delaunay N, Pichon V. Recent Advances in Lectin-Based Affinity Sorbents for Protein Glycosylation Studies. Front Mol Biosci. 2021;8.
48. Riley NM, Bertozzi CR, Pitteri SJ. A pragmatic guide to enrichment strategies for mass spectrometry–based glycoproteomics. Molecular & Cellular Proteomics. 2021;20.
49. Yamashita M, Fenn JB. Electrospray ion source. Another variation on the free-jet theme. J Phys Chem. 1984;88(20):4451–4459.
50. Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB. Electrospray interface for liquid chromatographs and mass spectrometers. Anal Chem. 1985;57(3):675–679.
51. Hillenkamp F, Karas M, Beavis RC, Chait BT. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal Chem. 1991;63(24):1193A-1203A.
52. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, Matsuo T. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid communications in mass spectrometry. 1988;2(8):151–153.
53. Fonslow BR, Yates Iii JR. Capillary electrophoresis applied to proteomic analysis. J Sep Sci. 2009;32(8):1175–1188.
54. Fenn JB. Electrospray wings for molecular elephants (Nobel lecture). Angewandte Chemie International Edition. 2003;42(33):3871–3894.
55. Cornett DS, Reyzer ML, Chaurand P, Caprioli RM. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. Nat Methods. 2007;4(10):828–833.
56. Zenobi R, Knochenmuss R. Ion formation in MALDI mass spectrometry. Mass Spectrom Rev. 1998;17(5):337–366.
57. Greco V, Piras C, Pieroni L, Ronci M, Putignani L, Roncada P, Urbani A. Applications of MALDI-TOF mass spectrometry in clinical proteomics. Expert Rev Proteomics. 2018;15(8):683–696.

58. Kruve A, Kaupmees K, Liigand J, Leito I. Negative electrospray ionization via deprotonation: predicting the ionization efficiency. Anal Chem. 2014;86(10):4822–4830.
59. Liigand P, Kaupmees K, Haav K, Liigand J, Leito I, Girod M, Antoine R, Kruve A. Think negative: finding the best electrospray ionization/MS mode for your analyte. Anal Chem. 2017;89(11):5665–5668.
60. Wilm M. Principles of electrospray ionization. Molecular & cellular proteomics. 2011;10(7).
61. Zhang Y, Fonslow BR, Shan B, Baek MC, Yates III JR. Protein analysis by shotgun/bottom-up proteomics. Chem Rev. 2013;113(4):2343–2394.
62. Garcia MC. The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography–electrospray mass spectrometry. Journal of Chromatography B. 2005;825(2):111–123.
63. Miller PE, Denton MB. The quadrupole mass filter: basic operating concepts. J Chem Educ. 1986;63(7):617.
64. Chernushevich I v, Loboda A v, Thomson BA. An introduction to quadrupole–time-of-flight mass spectrometry. Journal of mass spectrometry. 2001;36(8):849–865.
65. Boesl U. Time-of-flight mass spectrometry: introduction to the basics. Mass Spectrom Rev. 2017;36(1):86–109.
66. Douglas DJ, Frank AJ, Mao D. Linear ion traps in mass spectrometry. Mass Spectrom Rev. 2005;24(1):1–29.
67. Nikolaev EN, Kostyukevich YI, Vladimirov GN. Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. Mass Spectrom Rev. 2016;35(2):219–258.
68. Zubarev RA, Makarov A. Orbitrap mass spectrometry. ACS Publications; 2013.
69. Jones AW, Cooper HJ. Dissociation techniques in mass spectrometry-based proteomics. Analyst. 2011;136(17):3419–3429.
70. Olsen J v, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods. 2007;4(9):709–712.

71. Mechref Y. Use of CID/ETD mass spectrometry to analyze glycopeptides. Curr Protoc Protein Sci. 2012;68(1):11–12.
72. Halim A, Westerlind U, Pett C, Schorlemer M, Ruetschi U, Brinkmalm G, Sihlbom C, Lengqvist J, Larson G, Nilsson J. Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC–MS/MS of glycopeptides. J Proteome Res. 2014;13(12):6024–6032.
73. Wuhrer M, Catalina MI, Deelder AM, Hokke CH. Glycoproteomics based on tandem mass spectrometry of glycopeptides. Journal of Chromatography B. 2007;849(1–2):115–128.
74. Vékey K, Ozohanics O, Tóth E, Jekő A, Révész Á, Krenyácz J, Drahos L. Fragmentation characteristics of glycopeptides. Int J Mass Spectrom. 2013;345:71–79.
75. 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.
76. Yu Q, Wang B, Chen Z, Urabe G, Glover MS, Shi X, Guo LW, Kent KC, Li L. Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide/glycoproteome characterization. J Am Soc Mass Spectrom. 2017;28(9):1751–1764.
77. Frese CK, Altelaar AFM, van den Toorn H, Nolting D, Griep-Raming J, Heck AJR, Mohammed S. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. Anal Chem. 2012;84(22):9668–9673.
78. Hinneburg H, Stavenhagen K, Schweiger-Hufnagel U, Pengelley S, Jabs W, Seeberger PH, Silva DV, Wuhrer M, Kolarich D. The art of destruction: optimizing collision energies in quadrupole-time of flight (Q-TOF) instruments for glycopeptide-based glycoproteomics. J Am Soc Mass Spectrom. 2016;27(3):507–519.
79. Mitulović G. New HPLC techniques for proteomics analysis: a short overview of latest developments. J Liq Chromatogr Relat Technol. 2015;38(3):390–403.
80. Shishkova E, Hebert AS, Coon JJ. Now, more than ever, proteomics needs better chromatography. Cell Syst. 2016;3(4):321–324.

81. Stadlmann J, Hudecz O, Krššáková G, Ctortecka C, van Raemdonck G, op de Beeck J, Desmet G, Penninger JM, Jacobs P, Mechtler K. Improved sensitivity in low-input proteomics using micropillar array-based chromatography. Anal Chem. 2019;91(22):14203–14207.
82. Wilson SR, Vehus T, Berg HS, Lundanes E. Nano-LC in proteomics: recent advances and approaches. Bioanalysis. 2015;7(14):1799–1815.
83. Aydoğan C, Beltekin B, Alharthi S, Ağca CA, Erdoğan İY. Nano-liquid chromatography with monolithic stationary phase based on naphthyl monomer for proteomics analysis. J Chromatogr A. 2023;463804.
84. Chen SY, Dong M, Yang G, Zhou Y, Clark DJ, Lih TM, Schnaubelt M, Liu Z, Zhang H. Glycans, glycosite, and intact glycopeptide analysis of N-linked glycoproteins using liquid handling systems. Anal Chem. 2019;92(2):1680–1686.
85. Cao L, Qu Y, Zhang Z, Wang Z, Prytkova I, Wu S. Intact glycopeptide characterization using mass spectrometry. Expert Rev Proteomics. 2016;13(5):513–522.
86. Gianazza E, Tremoli E, Banfi C. The selected reaction monitoring/multiple reaction monitoring-based mass spectrometry approach for the accurate quantitation of proteins: clinical applications in the cardiovascular diseases. Expert Rev Proteomics. 2014;11(6):771–788.
87. Gallien S, Bourmaud A, Kim SY, Domon B. Technical considerations for large-scale parallel reaction monitoring analysis. J Proteomics. 2014;100:147–159.
88. Liu H, Sadygov RG, Yates JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004;76(14):4193–4201.
89. Zhang F, Ge W, Ruan G, Cai X, Guo T. Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. Proteomics. 2020;20(17–18):1900276.
90. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat Methods. 2004;1(1):39–45.
91. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent

acquisition: a new concept for consistent and accurate proteome analysis. Molecular & Cellular Proteomics. 2012;11(6).

92. Moseley MA, Hughes CJ, Juvvadi PR, Soderblom EJ, Lennon S, Perkins SR, Thompson JW, Steinbach WJ, Geromanos SJ, Wildgoose J. Scanning quadrupole data-independent acquisition, part A: qualitative and quantitative characterization. J Proteome Res. 2018;17(2):770–779.
93. Meier F, Brunner AD, Frank M, Ha A, Bludau I, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Bache N. diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. Nat Methods. 2020;17(12):1229–1236.
94. Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham AJL, Bunk DM, Kilpatrick LE, Billheimer DD, Blackman RK, Cardasis HL. Repeatability and reproducibility in proteomic identifications by liquid chromatography− tandem mass spectrometry. J Proteome Res. 2010;9(2):761–776.
95. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. ELECTROPHORESIS: An International Journal. 1999;20(18):3551–3567.
96. Clauser KR, Baker P, Burlingame AL. Role of accurate mass measurement (±10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal Chem. 1999;71(14):2871–2882.
97. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007;4(3):207–214.
98. Kawahara R, Alagesan K, Bern M, Cao W, Chalkley RJ, Cheng K, Choo MS, Edwards N, Goldman R, Hoffmann M. Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies of glycopeptide data. bioRxiv. 2021;2021–2023.
99. Cao W, Liu M, Kong S, Wu M, Zhang Y, Yang P. Recent advances in software tools for more generic and precise intact glycopeptide analysis. Molecular & Cellular Proteomics. 2021;20.

100. Wuhrer M, Catalina MI, Deelder AM, Hokke CH. Glycoproteomics based on tandem mass spectrometry of glycopeptides. Journal of Chromatography B. 2007;849(1–2):115–128.
101. Liu MQ, Zeng WF, Fang P, Cao WQ, Liu C, Yan GQ, Zhang Y, Peng C, Wu JQ, Zhang XJ. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. Nat Commun. 2017;8(1):438.
102. Kito K, Ito T. Mass spectrometry-based approaches toward absolute quantitative proteomics. Curr Genomics. 2008;9(4):263–274.
103. Urban PL. Quantitative mass spectrometry: an overview. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016;374(2079):20150382.
104. Ankney JA, Muneer A, Chen X. Relative and absolute quantitation in mass spectrometry–based proteomics. Annual review of analytical chemistry. 2018;11:49–77.
105. Tonry C, Finn S, Armstrong J, Pennington SR. Clinical proteomics for prostate cancer: understanding prostate cancer pathology and protein biomarkers for improved disease management. Clin Proteomics. 2020;17:1–31.
106. Viste E, Vinje CA, Lid TG, Skeie S, Evjen-Olsen Ø, Nordström T, Thorsen O, Gilje B, Janssen EAM, Kjosavik SR. Effects of replacing PSA with Stockholm3 for diagnosis of clinically significant prostate cancer in a healthcare system–the Stavanger experience. Scand J Prim Health Care. 2020;38(3):315–322.
107. Wu JT, Liu GH. Advantages of replacing the total PSA assay with the assay for PSA-α1-antichymotrypsin complex for the screening and management of prostate cancer. J Clin Lab Anal. 1998;12(1):32–40.
108. Shipitsin M, Small C, Choudhury S, Giladi E, Friedlander S, Nardone J, Hussain S, Hurley AD, Ernst C, Huang YE. Identification of proteomic biomarkers predicting prostate cancer aggressiveness and lethality despite biopsy-sampling error. Br J Cancer. 2014;111(6):1201–1212.
109. Duffy MJ. Biomarkers for prostate cancer: prostate-specific antigen and beyond. Clinical Chemistry and Laboratory Medicine (CCLM). 2020;58(3):326–339.

110. Tanase CP, Codrici E, Popescu ID, Mihai S, Enciu AM, Necula LG, Preda A, Ismail G, Albulescu R. Prostate cancer proteomics: Current trends and future perspectives for biomarker discovery. Oncotarget. 2017;8(11):18497.
111. Pin E, Fredolini C, Petricoin III EF. The role of proteomics in prostate cancer research: biomarker discovery and validation. Clin Biochem. 2013;46(6):524–538.
112. Gilgunn S, Conroy PJ, Saldova R, Rudd PM, O'kennedy RJ. Aberrant PSA glycosylation—a sweet predictor of prostate cancer. Nat Rev Urol. 2013;10(2):99–107.
113. Llop E, Ferrer-Batalle M, Barrabés S, Guerrero PE, Ramirez M, Saldova R, Rudd PM, Aleixandre RN, Comet J, de Llorens R. Improvement of prostate cancer diagnosis by detecting PSA glycosylation-specific changes. Theranostics. 2016;6(8):1190.
114. Nuhn P, de Bono JS, Fizazi K, Freedland SJ, Grilli M, Kantoff PW, Sonpavde G, Sternberg CN, Yegnasubramanian S, Antonarakis ES. Update on systemic prostate cancer therapies: management of metastatic castration-resistant prostate cancer in the era of precision oncology. Eur Urol. 2019;75(1):88–99.
115. Claps M, Mennitto A, Guadalupi V, Sepe P, Stellato M, Zattarin E, Gillessen SS, Sternberg CN, Berruti A, de Braud FGM. Immune-checkpoint inhibitors and metastatic prostate cancer therapy: Learning by making mistakes. Cancer Treat Rev. 2020;88:102057.
116. Seijo LM, Peled N, Ajona D, Boeri M, Field JK, Sozzi G, Pio R, Zulueta JJ, Spira A, Massion PP. Biomarkers in lung cancer screening: achievements, promises, and challenges. Journal of Thoracic Oncology. 2019;14(3):343–357.
117. Martínez-Terroba E, Behrens C, de Miguel FJ, Agorreta J, Monsó E, Millares L, Sainz C, Mesa-Guzman M, Pérez-Gracia JL, Lozano MD. A novel protein-based prognostic signature improves risk stratification to guide clinical management in early-stage lung adenocarcinoma patients. J Pathol. 2018;245(4):421–432.
118. Lehtiö J, Arslan T, Siavelis I, Pan Y, Socciarelli F, Berkovska O, Umer HM, Mermelekas G, Pirmoradian M, Jönsson M. Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune-evasion mechanisms. Nat Cancer. 2021;2(11):1224–1242.

119. Wadowska K, Bil-Lula I, Trembecki Ł, Śliwińska-Mossoń M. Genetic markers in lung cancer diagnosis: A review. Int J Mol Sci. 2020;21(13):4569.
120. Baran K, Brzeziańska-Lasota E. Proteomic biomarkers of non-small cell lung cancer patients. Adv Respir Med. 2021;89(4):419–426.
121. Gazdar AF, Gao B, Minna JD. Lung cancer cell lines: Useless artifacts or invaluable tools for medical science? Lung cancer. 2010;68(3):309–318.
122. Herath S, Sadeghi Rad H, Radfar P, Ladwa R, Warkiani M, O'Byrne K, Kulasinghe A. The role of circulating biomarkers in lung cancer. Front Oncol. 2022;11:801269.
123. Szasz AM, Győrffy B, Marko-Varga G. Cancer heterogeneity determined by functional proteomics. In: Seminars in Cell & Developmental Biology. Elsevier; 2017. p. 132–142.
124. Villalobos P, Wistuba II. Lung cancer biomarkers. Hematology/Oncology Clinics. 2017;31(1):13–29.
125. Fowler CB, O'Leary TJ, Mason JT. Toward improving the proteomic analysis of formalin-fixed, paraffin-embedded tissue. Expert Rev Proteomics. 2013;10(4):389–400.
126. Mantsiou A, Makridakis M, Fasoulakis K, Katafigiotis I, Constantinides CA, Zoidakis J, Roubelakis MG, Vlahou A, Lygirou V. Proteomics analysis of formalin fixed paraffin embedded tissues in the investigation of prostate cancer. J Proteome Res. 2019;19(7):2631–2642.
127. Negi A, Puri A, Gupta R, Chauhan I, Nangia R, Sachdeva A. Biosafe alternative to xylene: A comparative study. J Oral Maxillofac Pathol. 2013;17(3):363.
128. Kalantari N, Bayani M, Ghaffari T. Deparaffinization of formalin-fixed paraffin-embedded tissue blocks using hot water instead of xylene. Anal Biochem. 2016;507:71–73.
129. Shi SR, Cote RJ, Taylor CR. Antigen retrieval techniques: current perspectives. Journal of Histochemistry & Cytochemistry. 2001;49(8):931–937.
130. Leong TYM, Leong ASY. How does antigen retrieval work? Adv Anat Pathol. 2007;14(2):129–131.
131. Tanca A, Pagnozzi D, Burrai GP, Polinas M, Uzzau S, Antuofermo E, Addis MF. Comparability of differential proteomics data generated from paired archival fresh-

frozen and formalin-fixed samples by GeLC–MS/MS and spectral counting. J Proteomics. 2012;77:561–576.

132. Magdeldin S, Yamamoto T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. Proteomics. 2012;12(7):1045–1058.
133. Sugár S, Tóth G, Bugyi F, Vékey K, Karászi K, Drahos L, Turiák L. Alterations in protein expression and site-specific N-glycosylation of prostate cancer tissues. Sci Rep. 2021;11(1):15886.
134. Turiák L, Ozohanics O, Tóth G, Ács A, Révész Á, Vékey K, Telekes A, Drahos L. High sensitivity proteomics of prostate cancer tissue microarrays to discriminate between healthy and cancerous tissue. J Proteomics. 2019;197:82–91.
135. Turiák L, Sugár S, Ács A, Tóth G, Gömöry Á, Telekes A, Vékey K, Drahos L. Site-specific N-glycosylation of HeLa cell glycoproteins. Sci Rep. 2019;9(1):1–11.
136. Sugár S, Bugyi F, Tóth G, Pápay J, Kovalszky I, Tornóczky T, Drahos L, Turiák L. Proteomic Analysis of Lung Cancer Types—A Pilot Study. Cancers (Basel). 2022;14(11):2629.
137. Brosch M, Yu L, Hubbard T, Choudhary J. Accurate and sensitive peptide identification with Mascot Percolator. J Proteome Res. 2009;8(6):3176–3181.
138. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367–1372.
139. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote) omics data. Nat Methods. 2016;13(9):731–740.
140. R Core Team R. R: A language and environment for statistical computing. 2013;
141. RStudio RsT. Integrated development environment for R. RStudio, PBC: Boston, MA, USA. 2020;
142. Ozohanics O, Turiák L, Puerta A, Vékey K, Drahos L. High-performance liquid chromatography coupled to mass spectrometry methodology for analyzing site-specific N-glycosylation patterns. J Chromatogr A. 2012;1259:200–212.

143. Mering C von, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003;31(1):258–261.
144. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545–15550.
145. Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. Data mining in proteomics: from standards to applications. 2011;291–303.
146. Figel S, H Gelman I. Focal adhesion kinase controls prostate cancer progression via intrinsic kinase and scaffolding functions. Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents). 2011;11(7):607–616.
147. Wang Y, Gratzke C, Tamalunas A, Rutz B, Ciotkowska A, Strittmatter F, Herlemann A, Janich S, Waidelich R, Liu C. Smooth muscle contraction and growth of stromal cells in the human prostate are both inhibited by the Src family kinase inhibitors, AZM475271 and PP2. Br J Pharmacol. 2016;173(23):3342–3358.
148. Wang X, Li S. Protein mislocalization: mechanisms, functions and clinical applications in cancer. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer. 2014;1846(1):13–25.
149. Liang P, Pardee AB. Analysing differential gene expression in cancer. Nat Rev Cancer. 2003;3(11):869–876.
150. Wang M, Zhao J, Zhang L, Wei F, Lian Y, Wu Y, Gong Z, Zhang S, Zhou J, Cao K. Role of tumor microenvironment in tumorigenesis. J Cancer. 2017;8(5):761.
151. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer. 2015;15(9):540–555.
152. Kawahara R, Recuero S, Srougi M, Leite KRM, Thaysen-Andersen M, Palmisano G. The complexity and dynamics of the tissue glycoproteome associated with prostate cancer progression. Molecular & Cellular Proteomics. 2021;20.

153. Drake RR, Powers TW, Jones EE, Bruner E, Mehta AS, Angel PM. MALDI mass spectrometry imaging of N-linked glycans in cancer tissues. Adv Cancer Res. 2017;134:85–116.
154. Zhang C, Yan L, Song H, Ma Z, Chen D, Yang F, Fang L, Li Z, Li K, Li D. Elevated serum sialic acid levels predict prostate cancer as well as bone metastases. J Cancer. 2019;10(2):449.
155. Scott E, Munkley J. Glycans as biomarkers in prostate cancer. Int J Mol Sci. 2019;20(6):1389.
156. Zhu YP, Wan FN, Shen YJ, Wang HK, Zhang GM, Ye DW. Reactive stroma component COL6A1 is upregulated in castration-resistant prostate cancer and promotes tumor growth. Oncotarget. 2015;6(16):14488.
157. Shah P, Wang X, Yang W, Eshghi ST, Sun S, Hoti N, Chen L, Yang S, Pasay J, Rubin A. Integrated proteomic and glycoproteomic analyses of prostate cancer cells reveal glycoprotein alteration in protein abundance and glycosylation. Molecular & Cellular Proteomics. 2015;14(10):2753–2763.
158. Muniyan S, Chaturvedi NK, Dwyer JG, LaGrange CA, Chaney WG, Lin MF. Human prostatic acid phosphatase: structure, function and regulation. Int J Mol Sci. 2013;14(5):10438–10464.
159. Tian Y, Bova GS, Zhang H. Quantitative glycoproteomic analysis of optimal cutting temperature-embedded frozen tissues identifying glycoproteins associated with aggressive prostate cancer. Anal Chem. 2011;83(18):7013–7019.
160. Kohli M, Williams K, Yao JL, Dennis RA, Huang J, Reeder J, Ricke WA. Thrombin expression in prostate: a novel finding. Cancer Invest. 2011;29(1):62–67.
161. Adams G, Rosenfeldt L, Frederick M, Kombrinck K, Monia BP, Revenko A, Palumbo JS. Thrombin and factor XII drive prostate tumor growth in vivo. American Society of Hematology Washington, DC; 2015.
162. Nierodzik ML, Karpatkin S. Thrombin induces tumor growth, metastasis, and angiogenesis: Evidence for a thrombin-regulated dormant tumor phenotype. Cancer Cell. 2006;10(5):355–362.
163. Gudelj I, Lauc G, Pezer M. Immunoglobulin G glycosylation in aging and diseases. Cell Immunol. 2018;333:65–79.

164. Xu Y, Chen B, Zheng S, Wen Y, Xu A, Xu K, Li B, Liu C. IgG silencing induces apoptosis and suppresses proliferation, migration and invasion in LNCaP prostate cancer cells. Cell Mol Biol Lett. 2016;21:1–10.
165. Kazuno S, Furukawa J, Shinohara Y, Murayama K, Fujime M, Ueno T, Fujimura T. Glycosylation status of serum immunoglobulin G in patients with prostate diseases. Cancer Med. 2016;5(6):1137–1146.
166. Papakonstantinou M. N-Glycosylation of IgG Immunoglobulin and its clinical significance. J Biomed. 2019;4:35–43.
167. Yang J, Song H, Chen L, Cao K, Zhang Y, Li Y, Hao X. Integrated analysis of microfibrillar-associated proteins reveals MFAP4 as a novel biomarker in human cancers. Epigenomics. 2019;11(1):5–21.
168. Guerrero PE, Duran A, Ortiz MR, Castro E, Garcia-Velasco A, Llop E, Peracaula R. Microfibril associated protein 4 (MFAP4) is a carrier of the tumor associated carbohydrate sialyl-Lewis x (sLex) in pancreatic adenocarcinoma. J Proteomics. 2021;231:104004.
169. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F. A pathology atlas of the human cancer transcriptome. Science (1979). 2017;357(6352):eaan2507.
170. Dong M, Lih TM, Chen SY, Cho KC, Eguez RV, Höti N, Zhou Y, Yang W, Mangold L, Chan DW. Urinary glycoproteins associated with aggressive prostate cancer. Theranostics. 2020;10(26):11892.
171. Gabriele C, Cantiello F, Nicastri A, Crocerossa F, Russo GI, Cicione A, Vartolomei MD, Ferro M, Morgia G, Lucarelli G. High-throughput detection of low abundance sialylated glycoproteins in human serum by TiO 2 enrichment and targeted LC-MS/MS analysis: Application to a prostate cancer sample set. Anal Bioanal Chem. 2019;411:755–763.
172. Xu JY, Zhang C, Wang X, Zhai L, Ma Y, Mao Y, Qian K, Sun C, Liu Z, Jiang S. Integrative proteomic characterization of human lung adenocarcinoma. Cell. 2020;182(1):245–261.
173. El-Khoury V, Schritz A, Kim SY, Lesur A, Sertamo K, Bernardin F, Petritis K, Pirrotte P, Selinsky C, Whiteaker JR. Identification of a blood-based protein biomarker panel for lung cancer detection. Cancers (Basel). 2020;12(6):1629.

174. Winkler J, Abisoye-Ogunniyan A, Metcalf KJ, Werb Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat Commun. 2020;11(1):5120.
175. Tanjore H, Kalluri R. The role of type IV collagen and basement membranes in cancer progression and metastasis. Am J Pathol. 2006;168(3):715–717.
176. Bourgot I, Primac I, Louis T, Noël A, Maquoi E. Reciprocal interplay between fibrillar collagens and collagen-binding integrins: implications in cancer progression and metastasis. Front Oncol. 2020;10:1488.
177. Hou L, Lin T, Wang Y, Liu B, Wang M. Collagen type 1 alpha 1 chain is a novel predictive biomarker of poor progression-free survival and chemoresistance in metastatic lung cancer. J Cancer. 2021;12(19):5723.
178. Willumsen N, Bager CL, Leeming DJ, Bay-Jensen AC, Karsdal MA. Nidogen-1 degraded by cathepsin S can be quantified in serum and is associated with non–small cell lung cancer. Neoplasia. 2017;19(4):271–278.
179. Willumsen N, Bager CL, Leeming DJ, Bay-Jensen AC, Karsdal MA. Nidogen-1 degraded by cathepsin S can be quantified in serum and is associated with non–small cell lung cancer. Neoplasia. 2017;19(4):271–278.
180. Jagroop R, Martin CJ, Moorehead RA. Nidogen 1 regulates proliferation and migration/invasion in murine claudin-low mammary tumor cells. Oncol Lett. 2021;21(1):1.
181. Moon YW, Rao G, Kim JJ, Shim HS, Park KS, An SS, Kim B, Steeg PS, Sarfaraz S, Changwoo Lee L. LAMC2 enhances the metastatic potential of lung adenocarcinoma. Cell Death Differ. 2015;22(8):1341–1352.
182. Teng Y, Wang Z, Ma L, Zhang L, Guo Y, Gu M, Wang Z, Wang Y, Yue W. Prognostic significance of circulating laminin gamma2 for early-stage non-small-cell lung cancer. Onco Targets Ther. 2016;4151–4162.
183. Wei J, Hu M, Huang K, Lin S, Du H. Roles of proteoglycans and glycosaminoglycans in cancer development and progression. Int J Mol Sci. 2020;21(17):5983.
184. Han L, Sheng B, Zeng Q, Yao W, Jiang Q. Correlation between MMP2 expression in lung cancer tissues and clinical parameters: a retrospective clinical analysis. BMC Pulm Med. 2020;20(1):1–9.

185. Barkovskaya A, Buffone Jr A, Žídek M, Weaver VM. Proteoglycans as mediators of cancer tissue mechanics. Front Cell Dev Biol. 2020;8:569377.
186. Su C yue, Li J quan, Zhang L ling, Wang H, Wang F hua, Tao Y wen, Wang Y qing, Guo Q ru, Li J jun, Liu Y. The biological functions and clinical applications of integrins in cancers. Front Pharmacol. 2020;11:579068.
187. Maziveyi M, Alahari SK. Cell matrix adhesions in cancer: The proteins that form the glue. Oncotarget. 2017;8(29):48471.
188. Nagano M, Hoshino D, Koshikawa N, Akizawa T, Seiki M. Turnover of focal adhesions and cancer cell migration. Int J Cell Biol. 2012;2012.
189. Lin S, Li Y, Wang D, Huang C, Marino D, Bollt O, Wu C, Taylor MD, Li W, DeNicola GM. Fascin promotes lung cancer growth and metastasis by enhancing glycolysis and PFKFB3 expression. Cancer Lett. 2021;518:230–242.
190. Zhao D, Zhang T, Hou XM, Ling XL. Knockdown of fascin-1 expression suppresses cell migration and invasion of non-small cell lung cancer by regulating the MAPK pathway. Biochem Biophys Res Commun. 2018;497(2):694–699.
191. Fahrmann JF, Grapov D, Phinney BS, Stroble C, DeFelice BC, Rom W, Gandara DR, Zhang Y, Fiehn O, Pass H. Proteomic profiling of lung adenocarcinoma indicates heightened DNA repair, antioxidant mechanisms and identifies LASP1 as a potential negative predictor of survival. Clin Proteomics. 2016;13(1):1–12.
192. Chang W, Tsai M, Kuo P, Hung J. Role of galectins in lung cancer. Oncol Lett. 2017;14(5):5077–5084.
193. Chung LY, Tang SJ, Sun GH, Chou TY, Yeh TS, Yu SL, Sun KH. Galectin-1 Promotes Lung Cancer Progression and Chemoresistance by Upregulating p38 MAPK, ERK, and Cyclooxygenase-2Galectin-1 Activates COX-2 in Lung Cancer. Clinical Cancer Research. 2012;18(15):4037–4047.
194. Koundouros N, Poulogiannis G. Reprogramming of fatty acid metabolism in cancer. Br J Cancer. 2020;122(1):4–22.
195. Hatzivassiliou G, Zhao F, Bauer DE, Andreadis C, Shaw AN, Dhanak D, Hingorani SR, Tuveson DA, Thompson CB. ATP citrate lyase inhibition can suppress tumor cell growth. Cancer Cell. 2005;8(4):311–321.
196. Shi H, Huang H, Pu J, Shi D, Ning Y, Dong Y, Han Y, Zarogoulidis P, Bai C. Decreased pretherapy serum apolipoprotein AI is associated with extent of

metastasis and poor prognosis of non-small-cell lung cancer. Onco Targets Ther. 2018;6995–7003.

197. Fang S, Dai Y, Mei Y, Yang M, Hu L, Yang H, Guan X, Li J. Clinical significance and biological role of cancer-derived type I collagen in lung and esophageal cancers. Thorac Cancer. 2019;10(2):277–88.
198. Yang L, Lu P, Yang X, Li K, Qu S. Annexin A3, a calcium-dependent phospholipid-binding protein: Implication in cancer. Front Mol Biosci. 2021;8:716415.
199. Wang W, Hu Z, Sun C, Yao H. The role of annexin A4 in cancer. Frontiers in Bioscience-Landmark. 2016;21(5):949–957.
200. Tsuna M, Kageyama SI, Fukuoka J, Kitano H, Doki Y, Tezuka H, Yasuda H. Significance of S100A4 as a prognostic marker of lung squamous cell carcinoma. Anticancer Res. 2009;29(7):2547–2554.
201. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. Genes Dev. 2018;32(19–20):1267–1284.
202. Jiang C, Huang T, Wang Y, Huang G, Wan X, Gu J. Immunoglobulin G expression in lung cancer and its effects on metastasis. PLoS One. 2014;9(5):e97359.
203. Klotz M, Blaes F, Funke D, Kalweit G, Schimrigk K, Huwer H. Shift in the IgG subclass distribution in patients with lung cancer. Lung Cancer. 1999;24(1):25–30.
204. Wun T, White RH. Epidemiology of cancer-related venous thromboembolism. Best Pract Res Clin Haematol. 2009;22(1):9–23.
205. Huang W, Wang S, Zhang H, Zhang B, Wang C. Prognostic significance of combined fibrinogen concentration and neutrophil-to-lymphocyte ratio in patients with resectable non-small cell lung cancer. Cancer Biol Med. 2018;15(1):88.
206. Wang H, Zhao J, Zhang M, Han L, Wang M, Xingde L. The combination of plasma fibrinogen and neutrophil lymphocyte ratio (F-NLR) is a predictive factor in patients with resectable non small cell lung cancer. J Cell Physiol. 2018;233(5):4216–4224.
207. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–674.

208. Bjornsti MA, Houghton PJ. Lost in translation: dysregulation of cap-dependent translation and cancer. Cancer Cell. 2004;5(6):519–523.
209. Pecoraro A, Pagano M, Russo G, Russo A. Ribosome biogenesis and cancer: overview on ribosomal proteins. Int J Mol Sci. 2021;22(11):5496.
210. Desterro J, Bak-Gordon P, Carmo-Fonseca M. Targeting mRNA processing as an anticancer strategy. Nat Rev Drug Discov. 2020;19(2):112–129.
211. Hnisz D, Schuijers J, Li CH, Young RA. Regulation and dysregulation of chromosome structure in cancer. Annu Rev Cancer Biol. 2018;2:21–40.
212. Mamdani H, Jalal SI. DNA repair in lung cancer: potential not yet reached. Vol. 5, Lung Cancer Management. Future Medicine; 2016. p. 5–8.
213. Chang L, Fang S, Gu W. The molecular mechanism of metabolic remodeling in lung cancer. J Cancer. 2020;11(6):1403.
214. Zhang X, Liu D, Gao Y, Lin C, An Q, Feng Y, Liu Y, Liu D, Luo H, Wang D. The biology and function of extracellular vesicles in cancer development. Front Cell Dev Biol. 2021;9:777441.
215. Saviana M, Romano G, Le P, Acunzo M, Nana-Sinkam P. Extracellular vesicles in lung cancer metastasis and their clinical applications. Cancers (Basel). 2021;13(22):5633.
216. Tzeng HT, Wang YC. Rab-mediated vesicle trafficking in cancer. J Biomed Sci. 2016;23(1):1–7.

9. Bibliography of the candidate's publications

9.1. Publications Related to the Dissertation

1.

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

IF (2021): 4.997

2.

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

IF (2021): 6.575

9.2. Publications Unrelated to the Dissertation

1.

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

2.

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

3.

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

4.

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

5.

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

6.

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

7.

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

8.

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

9.

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

10.

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

11.

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

12:
Sugár, S., Drahos, L., & Vekey, K. (2023). Quantitative proteomics I.: Concept, design, and planning of quantitative proteomics experiments. Journal of Mass Spectrometry, 58(4), e4907.
13:
Pál, D., Tóth, G., Sugár, S., Fügedi, K. D., Szabó, D., Kovalszky, I., Papp, D., Schlosser, G., Tóth, C., Tornóczky, T., Drahos, L., & Turiák, L. (2023). Compositional analysis of glycosaminoglycans in different lung cancer types—a pilot study. International Journal of Molecular Sciences, 24(8), 7050.
14:
Balbisi, M., Sugár, S., Schlosser, G., Szeitz, B., Fillinger, J., Moldvay, J., Drahos, L., Szász, A. M., Tóth, G., & Turiák, L. (2023). Inter-and intratumoral proteomics and glycosaminoglycan characterization of ALK rearranged lung adenocarcinoma tissues: a pilot study. Scientific Reports, 13(1), 6268.
15:
Abruscato, G., Chiarelli, R., Lazzara, V., Punginelli, D., Sugár, S., Mauro, M., Librizzi, M., Di Stefano, V., Arizza V., Vizzini, A., Vazzana, M., & Luparello, C. (2023). In Vitro Cytotoxic Effect of Aqueous Extracts from Leaves and Rhizomes of the Seagrass Posidonia oceanica (L.) Delile on HepG2 Liver Cancer Cells: Focus on Autophagy and Apoptosis. Biology, 12(4), 616.

DOI:10.14753/SE.2023.2877

10. Acknowledgements

First, I would like to sincerely thank my supervisor Lilla Turiák for everything she has done for me during my PhD and before and highlight the inspiring and supportive environment she has created for us students to thrive in, and for being an inspiration and example throughout.

Second, I would like to thank everyone in general who helped and influenced me during my many years at the MS Proteomics Research Group in any way—directly or indirectly, professionally or personally (Kálmán Újszászy, László Drahos, Eszter Tóth, Olivér Ozohanics, Katalin Borsos, András Ács, Lilla Turiák, Károly Vékey, Gábor Tóth, Ágnes Révész, Márton Milley, Ágnes Gömöry, Fanni Bugyi, Dániel Szabó, Balázs Molnár, Domonkos Pál, Kata Fügedi, Kinga Nagy, Mirjam Balbisi).
I would like to say special thanks to Kálmán Újszászy, who was my first mentor and the one that introduced me to mass spectrometry and science in general; to László Drahos, who accepted me to the group and continuously supported me in many ways throughout; to Károly Vékey, who I have learnt a lot from regarding countless aspects of scientific research; and to Gábor Tóth, who has always inspired me to grow both professionally and personally.

Third, I would like to thank everyone outside of my workplace that supported me, from my parents—who always supported me no matter the circumstances—and my family, to the friends that shaped me into who I am today.

Finally, I would like to thank the two institutions, the Research Centre for Natural Sciences, and the Doctoral School of Pharmaceutical Sciences of Semmelweis University.