SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Analysis of Plasma C9 Epitope Heterogeneity

by Ilona Tornyi

Supervisor: Prof. Dr. László Takács

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR CELLULAR AND IMMUNE BIOLOGY

DEBRECEN, 2024

Analysis of Plasma C9 Epitope Heterogeneity

By Ilona Tornyi
Molecular biology MSc

Supervisor: Prof. Dr. László Takács

Doctoral School of Molecular Cell and Immune Biology, University of Debrecen

Head of the **Defense Committee**:	Prof. Dr. László Fésüs, DSc, MHAS
Reviewers:	Prof. Dr. László Prókai, DSc
	Dr. Zsolt Megyesfalvi, PhD
Members of the Defense Committee:	Prof. Dr. Éva Csősz, DSc
	Prof. Dr. Balázs Sarkadi, DSc

The PhD Defense takes place at the Lecture Hall of Bldg. A, Department of Internal Medicine, Faculty of Medicine, University of Debrecen, at 2 p.m. on 3rd of February, 2025.

1. Introduction and Literature Review

Lung cancer has the highest mortality rate among cancers, with millions of deaths worldwide each year. The five-year survival rate ranges between 10-20%. According to statistics, Hungary had the highest rate of lung cancer deaths in Europe in 2019. The survival of patients is influenced by diagnosing the disease at the earliest possible stage. Based on the results of several large international screening programs, LDCT imaging is currently the recommended method for early-stage screening. Although these programs have detected more early-stage lung cancer patients, the number of false-positive results was high in all studies. Various methods are being tried to reduce this high rate, such as other imaging techniques (magnetic resonance imaging), risk factor-based algorithms, artificial intelligence, or biomarkers. With the advancement of biotechnology, increasingly sophisticated methods are available for developing and validating biomarkers.

A new direction initiated by our research group, epitope mapping, involves approaching proteins through their epitopes. One strategy of this is epitope profiling, which entails examining epitopes of proteins at the sub-protein level and identifying proteoforms that can be distinguished through their epitopes. This approach can significantly increase the number of known and examinable proteoforms derived from a single gene, as the number of proteins presumed based on genome sequence is not around 20,000, but much higher due to variability. The variation in proteins can affect their structure, thereby influencing the structure and accessibility of epitopes.

Using the epitope profiling toolkit previously developed by our group, including the PlasmaScan and QantiPlasma monoclonal libraries, it is possible to examine multiple epitopes of a single protein, thus detecting proteoforms. Different proteoforms may be associated with various pathological conditions, thereby providing biomarker functions to the epitopes that detect these proteoforms.

In the literature review section of my thesis, I will examine the epidemiology of lung cancer, its forms, and possible causes of development, as well as the results of screening studies conducted so far. The results of international LDCT screening studies have shown that CT-based screening needs to be supplemented with other methods to achieve better results. Such a complementary method could be biomarker analysis with a new approach of epitope profiling. During my work, I examined three epitopes of the complement C9 protein with monoclonal antibodies in control and lung cancer samples. With our results, we aim to contribute to the

development of new epitope-based plasma protein biomarkers (such as C9 biomarkers), which could successfully complement current lung cancer screening programs

1.1. Epidemiology of Lung Cancer Worldwide

Isaac Adler wrote the first English monograph entirely dedicated to lung cancer in 1912, in which he collected 374 case studies. Although this number of cases seems high, it is considered low given that by 1912, Adler found only this many in the literature, which regarded lung cancer as a very rare disease. Adler was the first to identify a link between smoking and lung cancer in his study.

Nowadays, cancer is the leading clinical, social, and economic burden worldwide. Lung cancer is the second most frequently diagnosed cancer and the leading cause of cancer mortality globally, considering both sexes. The average five-year survival rate is around 10-20% for those diagnosed with lung cancer between 2010-2014. This rate is higher than average in Japan (33%), Israel (27%), and Korea (25%).

In 2019, 55% of deaths could be attributed to 10 leading causes. These 10 causes led to the deaths of 55.4 million people worldwide. Of these causes, seven were chronic diseases, accounting for 74% of all deaths. Among these, as the sixth leading cause, and the only type of cancer, were malignant neoplasms of the bronchus and lung (ICD-10: C34). This type of cancer claimed 1.2 million lives in 2000 and 1.8 million in 2019.

1.2. Epidemiology of Lung Cancer in Hungary

In Europe, Hungary had the highest rate of cancer deaths (320.87 per 100,000 population) in 2020. The European Union average was 242.24 per 100,000 population in the same year. The lung cancer mortality rate was the highest in Hungary among European countries in 2019 (84.4 per 100,000 population). A summary publication in 2018, which examined the 25 most common cancer types in 40 European countries, showed that Hungary had the highest lung cancer incidence (111.6 per 100,000 men; 58.7 per 100,000 women) and lung cancer mortality rate (97.1 per 100,000 men; 44.3 per 100,000 women). A 2019 Hungarian study, using data from the National Health Insurance Fund (NEAK), calculated lung cancer mortality rates for the period between 2011 and 2016. The study included patients who: i) were newly diagnosed with lung cancer between January 1, 2011, and December 31, 2016; ii) appeared for examination twice within 30 days to one year; iii) were over 20 years old at diagnosis; iv) if the patient died

within 60 days after diagnosis. Exclusion criteria were: i) if another cancer was diagnosed within six months or after 12 months following the lung cancer diagnosis; ii) if the patient received other cancer treatments within six months or after 12 months of lung cancer treatment. Considering these criteria, 7158 new patients were registered in 2011 and 6996 in 2016 in the National Health Insurance Fund database. The Hungarian study concluded: i) although the age-standardized incidence rate is high in the country, it is lower than what was shown in Ferlay's 2018 review; ii) the age-standardized incidence rate shows a decreasing trend in men and an increasing trend in women. In the Ferlay study, statistical models calculated the estimated incidence and mortality rates from national cancer registry data, if available, or estimated them otherwise. Since Hungary did not report national incidence data during the period examined in the Ferlay study, it was estimated using WHO data from neighboring countries' incidence and mortality rates. Additionally, only mortality data were available from the Hungarian Central Statistical Office (KSH), derived from postmortem examinations. Hungary has a very high number of autopsies following hospital deaths. Higher cancer incidence rates can be detected during autopsies compared to clinical findings. The differences between the two studies can be explained by differences in data collection and the high rate of autopsies following hospital deaths. In Hungary, during the period between 2011-2016, the lung cancer mortality rate was high (72.8 per 100,000 men; 28.2 per 100,000 women), but likely not as high as international statistics indicated (96.4 per 100,000 men; 37.7 per 100,000 women).

1.3. Histological Classification of Lung Cancer

According to WHO guidelines, lung cancer is divided into two main histological groups: non-small cell lung cancer (NSCLC), which occurs in 85% of cases, and small cell lung cancer (SCLC), accounting for 15% of cases. NSCLC can be further subdivided into adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and other types. The progression of the disease varies among individuals, with the growth rate of NSCLC ranging from very fast to very slow, and in some cases, regression can occur. Adenocarcinoma grows more slowly than squamous cell carcinoma. Poorly differentiated tumors grow faster than well-differentiated ones. The growth rate of tumors varies by stage: it is faster in the early stage, slows or stagnates in stages II and III, and accelerates again in stage IV. The TNM classification system recommended by the TNM atlas is used for tumor staging and treatment. Lung cancer is a complex disease influenced by both genetic and environmental factors. Understanding the biology of lung cancer is essential for selecting the appropriate treatment.

1.4. Causes of Lung Cancer

1.4.1. Genetic Causes

The factors contributing to the development of lung cancer can be divided into three groups: (i) single-gene Mendelian inherited defects predisposing to lung cancer, (ii) polygenic polymorphisms predisposing to lung cancer, and (iii) mutations occurring in the tumor that are directly responsible for tumor growth.

i. Predisposing single-gene defects: mutations in DNA repair and other genes can rarely contribute to the development of lung cancer.

ii. Predisposing polygenic polymorphisms: polymorphisms of the nicotinic acetylcholine receptor in the 15p25-24 region (CHRNA5, CHRNA3, and CHRNB4), along with variants in other genes, can be sensitizing loci for lung cancer. The combined occurrence of hereditary variants associated with lung cancer is responsible for the familial aggregation of lung cancer.

iii. Tumor-associated mutations directly responsible for tumor growth: oncogenes include receptor tyrosine kinases such as EGFR, ALK, ROS1, and RET, as well as genes affected by gene amplification like MET and FGFR1. Other tyrosine kinases, such as KRAS and BRAF, are also common. Tumor suppressor genes include TP53, the most frequently mutated gene in lung cancer, as well as STK11, CDKN2A, KEAP1, and SMARCA4. According to the histological classification of driver mutations, EGFR, KRAS, ALK, c-met, AKT/PI3K, PTEN, ROS-1, and BRAF mutations are common in adenocarcinomas. In squamous cell carcinomas, c-met, VEGFR, and P53/BCL mutations are common, while in large cell and small cell lung cancer, c-met mutations also dominate. These mutations influence tumor survival and growth through different signaling pathways, complicating the uniform treatment of subtypes classified under the same category. A cancer cell may harbor one to ten critically defining "driver" mutations that support the cell's survival.

1.4.2. Environmental Factors

The development of lung cancer is influenced by various environmental factors, including physical and chemical agents, as well as smoking. Physical factors include natural background radiation, particularly radon, which is the leading cause of lung cancer among non-smokers according to the United States Environmental Protection Agency. Air pollution, such as particulate matter measuring 10 and 2.5 microns, also increases the risk of lung cancer, especially with long-term exposure. Exhaust from diesel vehicles has also been classified as carcinogenic. Chemical factors include chloramphenicol, dioxin, and phenoxy acid derivatives found in pesticides, as well as asbestos, which can cause lung cancer and other malignancies. Smoking is the most significant risk factor for lung cancer, as the compounds found in cigarettes are carcinogenic. Reducing smoking could significantly decrease lung cancer mortality rates. Although the relationship between marijuana uses and lung cancer is not clear, some studies suggest that THC can trigger malignant cell proliferation. Smoking is a particularly severe public health issue because it also causes other serious diseases, such as cardiovascular diseases and COPD. Passive smoking and air pollution are also significant risk factors. Prevention efforts must focus on public education and regular screening examinations.

1.5. Lung Cancer Screening

Early diagnosis of lung cancer can significantly improve survival rates for patients. For non-small cell lung cancer, the 5-year survival rate is 65% in the localized phase, but only 9% with distant metastases. For small cell lung cancer, these rates are 30% and 3% with metastases, respectively. Early detection is crucial as surgical, chemo-radio, and immunotherapeutic treatments can be effective in early stages, while the main goal in late-stage treatment is to reduce tumor size and alleviate symptoms.

Screening is critical in reducing lung cancer mortality since treatment outcomes are significantly better for early-stage lung cancer. Chest X-rays are not the most effective method for lung cancer screening. Currently, international guidelines recommend LDCT scans as the best method for detecting early-stage lung cancer, although this method also involves some radiation exposure.

1.5.1. International Screening Programs

The results of major international screening programs aimed at early detection of lung cancer show significant progress. In the American National Lung Screening Trial (NLST) (54,454 participants, with 26,732 undergoing X-ray and 26,722 undergoing LDCT), research found a 20% reduction in mortality in the LDCT group compared to the chest X-ray group. In the Dutch-Belgian NELSON trial (7,557 participants screened with LDCT and 8,265 in the control group), following nodule volume reduced false-positive results to 64.3%; mortality decreased by 24% in men and 33% in women. In the Italian DANTE trial (1,276 participants screened with LDCT, 1,196 in the control group), lung cancer was found in 4.7% of the LDCT group and 2.8% of the control group. In the ITALUNG study (1,406 participants screened with LDCT, 1,593 in the control group), 30.3% of the LDCT group were LC-positive, and 21 cases of lung cancer were diagnosed, nearly half of which were in stage I. In the Danish DLCST trial (2,052 participants screened with LDCT, 2,052 in the control group), 17 cases of lung cancer were found, more than half of which were in stage I. In the UKLS trial in England (1,994 participants screened with LDCT, 2,027 in the control group), 42 cases of lung cancer were diagnosed in the LDCT group, two-thirds of which were in stage I. In the German LUSI study (2,029 participants screened with LDCT, 2,023 in the control group), LDCT screening reduced mortality in women but showed no significant difference in men. It can be generally concluded that LDCT screening is more effective than chest X-rays in early detection of lung cancer and reducing mortality.

1.5.2. The HUNCHEST Program

Following the first major lung cancer screening programs, national screening programs were launched in Hungary as well. In the HUNCHEST I program (2014-2018), 739 participants were screened at the National Korányi Institute of Pulmonology, where malignant lesions were found in 38 participants, and 26 cases were false positives. The program's results align with international studies. The HUNCHEST II program (2019-2022) was Hungary's largest LDCT screening program, selecting 4,215 high-risk individuals. During the examination, 76 lung cancer patients were diagnosed, 62 of them in the first screening round. Most malignant tumors detected by screening were in stage I, and only 16.4% were in stage IV.

1.6. Efforts to Reduce High False-Positive Results

It is crucial to reduce the number of high false-positive results in LDCT screenings to alleviate participants' fears and increase cost-effectiveness. Identifying risk groups based on various risk factors can help narrow down the target population for screening. More advanced CT equipment and artificial intelligence can contribute to more accurate analysis of results. While there are attempts to use other imaging methods, such as MRI, because it does not involve radiation exposure, studies have shown that the sensitivity of MRI approaches that of LDCT for nodules measuring 8-10 mm. Additionally, liquid biopsy tests can complement LDCT examinations. These tests are minimally invasive, do not involve additional radiation exposure, and blood sampling can be performed in a general practitioner's office. These tests provide information about the patient's physical condition using biomarkers.

1.7. Biomarkers

Biomarkers are substances, structures, or processes that indicate normal or abnormal biological processes, therapeutic responses, or the presence of diseases. They can be classified based on proteins, nucleic acids, metabolites, and other molecules, and can be measured from various body fluids or other samples such as blood, urine, saliva, exhaled air, and stool. Their use spans a wide range of applications, including screening, diagnosis, prognosis, personalized therapy, and monitoring therapeutic effectiveness. The development of biomarkers occurs in five phases: discovery, qualification, verification, validation, and clinical use. The ideal cancer biomarker possesses high sensitivity, specificity, positive and negative predictive values, and, when used for screening, is capable of detecting tumors at an early stage. Additionally, in other applications, it provides prognostic and therapeutic values. With the advancements in modern molecular biology and immunology methods, biomarkers are playing an increasingly significant role in clinical diagnostics and therapeutic decision-making. Nowadays, most biomarkers are single or multi-platform based multivariate disease risk determinants.

Lung cancer biomarkers can come from various biological samples, such as blood, bronchoalveolar lavage, saliva, pleural fluid, and exhaled air. The most commonly used biomarkers include protein-based and nucleic acid-based biomarkers. Protein-based biomarkers include CYFRA21-1, CEA, SCC, NSE, ProGRP, and EGFR. Some of these are specific to lung cancer types, while others are found in other tumors as well. For instance, CYFRA21-1 and CEA are used in non-small cell lung cancer, whereas NSE and ProGRP are applicable in small cell lung cancer. Additionally, genomic markers (cfDNA, miRNA, epigenetic markers) are also used. The greatest hopes were associated with the GRAIL/Illumina Galleri test, a multi-cancer

liquid biopsy that focuses on known tumor mutations by examining circulating free DNA. Metabolite biomarkers (e.g., palmitic acid, heptadecanoic acid) could also be potential biomarkers that aid in the early diagnosis and treatment of lung cancer.

1.8. The Proteome

Proteins perform their functions in cells in their three-dimensional structures, which can be described at four levels: primary, secondary, tertiary, and quaternary structures. The primary structure is the linear sequence of amino acids in a protein. The secondary structure includes helices, sheets, and loops. Tertiary structures show the three-dimensional arrangement of peptide chains, while quaternary structures describe the connection of multiple protein chains. The concept of the proteome extends beyond genetic information to include the diversity of proteoforms, which can be much greater in number than the genes encoding human proteins. The variability of the proteome is influenced by post-translational modifications, alternative translational starts, termination sites, splicing, and other factors that vary from gene to gene.

1.9. Epitopes

Protein epitopes are three-dimensional arrangements of amino acids on the surface of proteins. These are recognized by components of the immune system (antibodies, T-cell receptors) through antigen-binding sites. Epitopes generally encompass 6-12 amino acids and bind to antibodies via weak interactions. The structure and accessibility of epitopes can vary due to changes in the tertiary and quaternary structure of the protein and changes in the composition of protein complexes. These changes can affect the functional relationships and biological roles of epitopes. Current proteomic technologies cannot globally track the variability and proteoforms of proteins but are important tools for studying protein diversity. The dynamic changes in protein epitopes can be associated with certain physiological or pathological conditions and may potentially serve as biomarkers or indicators of functional changes.

1.10. The Complement Cascade

The complement system is a complex part of innate immunity, natural immunity, consisting of more than 40 soluble proteins. The complement system performs various functions, such as clearing certain pathogenic bacteria, virus-infected and apoptotic cells, maintaining tissue homeostasis, and regulating the innate immune response. The system can be activated through

three pathways: the classical, lectin, and alternative pathways. During activation, C3 convertase is formed, which cleaves C3 into C3a and C3b. C3b plays a role in the formation of C5 convertase, which leads to the cleavage of C5 into C5a and C5b fragments. C5b forms a complex with C6, C7, and C8 proteins, embedding into the membrane and attracting C9 proteins, forming the membrane attack complex (MAC). The C9 proteins create a ring-shaped hole in the membrane of the pathogenic bacteria or, the virus-infected cell, thereby playing a role in the destruction of the pathogen.

The C9 proteins form a complex of 22 symmetric monomers, binding to the C5b678 complex and capable of creating a ring-shaped pore with a diameter of 120 Å in the pathogen membrane. The C9 protein has four main domains: the N-terminal thrombospondin-1 (TSP1) domain, low-density lipoprotein receptor-associated domain (LDLRA), membrane attack complex/perforin (MACPF) domain, and epidermal growth factor (EGF) domain. Since C9 does not contain a specific membrane-binding domain, it circulates in the plasma and associates with other parts of the MAC. The C9 monomers connect through the MACPF domain. The formation of an unintentional MAC is prevented by extracellular chaperones (vitronectin, clusterin) binding to soluble MAC complexes. C9 can be produced in the liver, monocytes, and dendritic cells.

1.11. Epitope Profiling

In everyday clinical practice, genome and transcriptome sequencing tests are already in use. However, global protein profiling technology is not yet available, despite the fact that understanding protein function through global profiling would fill a significant gap. Proteins, as the most common type of biomarker, are widely used to examine physiological abnormalities. While protein studies can also provide information for detecting abnormalities, examining epitopes within protein units offers a more sensitive solution. Each variant and variation of a protein can affect its tertiary and quaternary structures and the accessibility of epitopes. Protein epitope profiling (PEP) involves measuring the dynamics of epitopes using a monoclonal antibody library (QuantiPlasma Library) and a single mAb-based competitive inhibition ELISA system (sbCIA).

Our previous results have shown that certain protein epitopes have biomarker value in lung cancer associations, and these associations are not necessarily uniform, as we observed positive, negative, and neutral associations. The QuantiPlasma mAb library used for PEP was generated with normalized human plasma (for concentration and immunogenicity) as the immunogen, ensuring that the mAbs recognize natural antigenic determinants. The hybridomas were selected

to identify mAbs that function in the sbCIA system, meaning the degree of binding shows a linear relationship with plasma dilution. After selecting the hybridomas, monoclonal antibodies can be produced in vitro and in vivo. Monoclonal antibodies react with specific epitopes, allowing multiple epitopes to be associated with a single protein.

We estimate that the PlasmaScan and QuantiPlasma libraries contain approximately three epitopes per protein for around 100 plasma proteins. PEP technology allows monitoring not only the quantity but also the qualitative changes of proteins under different conditions. Unlike current global proteomic methods, PEP focuses on the vast array of epitopes at the subprotein level. In a previous publication, the association of leucine-rich alpha-2-glycoprotein 1 (LRG1) protein with lung cancer was detected using the sandwich ELISA method, revealing that certain accessible epitopes of LRG1 immune complexes do not uniformly associate with lung cancer. This observation highlighted the potential for broader subprotein, epitope-based investigations. We examined epitopes and their association with cancer using the fraction of the PlasmaScan library that functions in the sbCIA assay (~380 mAb), known as the QuantiPlasma library, during PEP. Our results indicated that, for the C9 molecule, certain anti-C9 specific mAbs and their recognized epitopes showed different associations with lung cancer (neutral [BSI0449], positive [BSI0639], and negative [BSI0581]). I aimed to investigate the molecular proteomic basis of this phenomenon.

In my work, I sought to understand the molecular background of the observed epitope heterogeneity phenomenon for a single protein through the epitopes of the C9 protein.

2. Aims

During my initial hypothesis-free research, we utilized C9 epitopes to search for and examine lung cancer-associated epitope biomarkers at the subprotein level, aiming to demonstrate the efficiency of PEP technology using C9 as an example. The objectives of my current research were as follows:

1. Statistical analysis of data from our previous research results. Selection of anti-C9 antibodies.
2. Purification and validation of the selected anti-C9 mAbs. Validation of the selected mAbs with commercially available C9 protein, including antigen recognition and technical validation in Western blot experiments.
3. Examination of the overlap of epitopes on C9, specifically investigating whether the studied epitopes are independent of each other.
4. Protein chemical analysis of C9 epitope-dependent heterogeneity:
    a. Using SDS PAGE method with blood plasma mixtures from lung cancer patients and control individuals,
    b. LC-MS/MS examination of the above samples through immunoprecipitation of individual epitopes.

The results of the initial hypothesis-free research led to further objectives:

5. Data analysis of LC-MS/MS results, primarily focusing on the structure of C9, N-glycosylation at position 415, and other plasma proteins associated with C9.
6. Building hypotheses regarding C9 proteoforms, the C9 interactome, and the subprotein level examination efficiency of PEP technology based on the results obtained from data analysis.

3. Materials and Methods

3.1. Examined Patient Groups and Processed Samples

The pooled control samples were obtained from 167 healthy donors, all of whom signed consent forms prior to sample collection. The sample collection was approved by the DE Regional and Institutional Research Ethics Committee (DE OEC RKEB/IKEB) under permits 3049-2009 and 3140-2010. The pooled patient samples were derived from 207 late-stage (Stage IV) lung cancer patients from four centers in Hungary. This sample collection was approved by the Scientific and Research Ethics Committee of the Scientific Council (ETT TUKEB) under permits 11739/2014/EKU, 107/2014, and 417/2014.

The study involved various centers across Hungary, including the Pulmonology Clinic of Semmelweis University, the Pulmonology Clinic of the University of Szeged, the Pulmonology Clinic of the University of Debrecen, and the Pulmonology Department of the Central Hospital of Borsod-Abaúj-Zemplén County.

During the study, we established the inclusion and exclusion criteria for lung cancer patients. The inclusion criteria included histologically confirmed lung cancer, supported by staging; participants were aged between 45 and 75 years, and had a smoking history of at least 15 pack-years. The exclusion criteria included any other cancer diagnosis within the past 5 years (except for non-metastatic skin cancer), active chronic disease, steroid treatment, as well as active tuberculosis and certain other active infectious diseases.

For the control group, healthy individuals, screening patients, as well as hospital staff and their relatives were recruited. To ensure the selected control group was truly appropriate, their suitability was confirmed through clinical examinations, LDCT, chest X-rays, biomarker analysis, and laboratory tests.

Collected EDTA samples were centrifuged at 4 °C, and the plasma was stored at -70 °C until use.

3.2. Immunoprecipitation

For immunoprecipitation, Dynabeads™ Protein G (Invitrogen, Carlsbad, CA, USA) magnetic beads were used. For each monoclonal antibody (mAb), 50 µl of Protein G was pre-washed with 200 µl of W&B (Wash and Binding) buffer (0.1 M Na2HPO4x12H2O [VWR, Lutherworth, Leicestershire, UK], 0.05% Tween [Sigma, St. Louis, MO, USA], pH 8.2). After

bead preparation, 75 µg of mAb, diluted in 200 µl of W&B buffer, was added to the pre-washed beads and incubated at room temperature on a mixer rotating at 5 RPM. The beads were then washed with 200 µl of PBS buffer (123 mM NaCl [Spektrum 3D, Debrecen, Hungary], 3.16 mM KH2PO4 [Riedel-de Haën, Seelze, Germany], 10.45 mM Na2HPO4x2H2O [Riedel-de Haën, Seelze, Germany]), followed by 200 µl of distilled water diluted 0.2 M triethanolamine. After the final wash, 200 µl of freshly prepared 20 mM DMP (Dimethyl Pimelimidate [Thermo Scientific, Waltham, MA, USA]) in 0.2 M triethanolamine diluted in distilled water was added to the magnetic beads and incubated at room temperature for 30 minutes on a mixer rotating at 5 RPM. The supernatants were removed, and 200 µl of 0.2 M distilled water-diluted monethanolamine (Sigma, St. Louis, MO, USA) was added to the beads, incubated for 15 minutes on a mixer rotating at 5 RPM. After removing the supernatant, the beads were washed three times with 200 µl of PBS buffer. Following antibody binding, 200 µl of plasma sample, diluted in 200 µl of PBS buffer and mixed with 100 µl of 20x protease inhibitor (1 SIGMAFAST™ Protease Inhibitor Tablet (Sigma, St. Louis, MO, USA) dissolved in 20 ml of distilled water), was added to the beads attached to the mAbs. This mixture was incubated overnight on a mixer at 4 °C. The next day, after centrifugation and removal of the supernatant, the beads were washed three times with 200 µl of PBS buffer. The beads were then transferred to clean Eppendorf tubes with 100 µl of PBS buffer. After removing the supernatant, 50 µl of 50 mM glycine (Sigma, St. Louis, MO, USA) solution in distilled water (pH 2.5) was added to the beads and incubated at room temperature for 10 minutes. The eluate was transferred to a clean Eppendorf tube containing 10 µl of 1 M Tris solution (Sigma, St. Louis, MO, USA; pH 9.0).

3.3. SDS-PAGE

Reducing gels were used for SDS-PAGE. To prepare the 12% separating gel, we used 4.9 ml of distilled water, 2.4 ml of Proseive (Lonza, Basel, Switzerland) solution, 2.5 ml of 1.5 M Tris solution (pH 8.8), 100 µl of 10% SDS (Sigma, St. Louis, MO, USA) solution, 100 µl of ammonium persulfate (APS [Sigma, St. Louis, MO, USA]), and 5 µl of tetramethylethylenediamine (TEMED [Thermo Scientific, Waltham, MA, USA]). The incubation time was 45 minutes. To prepare the 5% stacking gel, we used 3.75 ml of distilled water, 500 µl of Proseive (Lonza, Basel, Switzerland) solution, 650 µl of 1 M Tris solution (pH 6.8), 50 µl of 10% SDS solution, 50 µl of APS, and 5 µl of TEMED. The samples were loaded

onto the gel in a 6x running buffer (pH 6.8) containing bromophenol blue dye after a 10-minute denaturation at 95 °C. The 10x7 cm gels were run in a running buffer (pH 8.4) containing SDS, glycine, and Tris for 1-1.5 hours at 130 V.

3.4. Staining

3.4.1. Coomassie Staining

After electrophoresis, the gels were stained for 1 hour in a staining tank with Coomassie Brilliant Blue G-250 (Sigma, St. Louis, MO, USA) staining solution. After staining, the gels were destained with a destaining solution (a 1:1 mixture of 50% ethanol and methanol [Scharlab, Debrecen, Hungary]).

3.4.2. Silver Staining

Following Coomassie staining, the gels were also silver-stained. The destained gels were fixed for 20 minutes in a solution containing 50% methanol and 5% acetic acid. They were then soaked for 20 minutes in a 50% methanol solution. After a 10-minute soak in ultrapure water, the gels were sensitized for 1 minute in a 1.27 mM $Na_2S_2O_3$ (Sigma, St. Louis, MO, USA) solution and washed twice for 1 minute in ultrapure water. The gels were stained for 20 minutes in a 0.1% $AgNO_3$ (Sigma, St. Louis, MO, USA) solution, then washed twice for 1 minute in ultrapure water. Development was carried out with intense shaking in a solution containing 2% Na_2CO_3 (Sigma, St. Louis, MO, USA) and 0.04% formaldehyde (Sigma, St. Louis, MO, USA) until a suitably discriminative color density was achieved. The reaction was stopped in 5% acetic acid for 3-5 minutes. The silver-stained gels were stored in a 1% acetic acid (Scharlab, Debrecen, Hungary) solution at 4 °C. Images of the stained gels were digitally recorded with a GelLogic 2200Pro (Carestream, Rochester, NY, USA) device.

3.5. Western Blot

Reducing gels were used for Western blotting. To prepare the 12% separating gel, we used 4.9 ml of distilled water, 2.4 ml of Proseive (Lonza, Basel, Switzerland) solution, 2.5 ml of 1.5 M Tris solution (pH 8.8), 100 µl of 10% SDS solution, 100 µl of APS, and 5 µl of TEMED. The

incubation time was 45 minutes. To prepare the 5% stacking gel, we used 3.75 ml of distilled water, 500 µl of Proseive (Lonza) solution, 650 µl of 1 M Tris solution (pH 6.8), 50 µl of 10% SDS solution, 50 µl of APS, and 5 µl of TEMED. The samples were loaded onto the prepared 10x7 cm gels with a 6x loading buffer containing mercaptoethanol and bromophenol blue after a 10-minute incubation at 95 °C. After approximately 1-1.5 hours of running at 130 V, the gels were placed into a blotting cassette. The layers within the blotting cassette were arranged as follows: 1. sponge, 2. filter paper, 3. membrane, 4. gel, 5. filter paper, 6. sponge. The protein transfer to the membrane was carried out for 1 hour at 150 V with ice cooling. After blotting, the membrane was blocked overnight at 4 °C on a horizontal shaker in blocking buffer (6.25 mM polyvinylpyrrolidone, 18.8 mM NaCl, 50 ml PBS-Tween). The next day, the membrane was incubated overnight at 4 °C with the primary antibody diluted in blocking buffer, then washed 5 times for 10 minutes in a PBS-Tween solution. The membrane was incubated with the secondary antibody for 1 hour at room temperature, then washed 5 times for 10 minutes in PBS-Tween solution. For detection, Pierce ECL Western Blotting Substrate (Thermo Scientific, Waltham, MA, USA) was used in a 1:1 ratio. Images were digitally recorded with a GelLogic 2200Pro (Carestream, Rochester, NY, USA) device.

3.6. Mass Spectrometry

3.6.1. Sample Preparation

Whole plasma samples were immunoprecipitated using C9-specific antibodies (BSI0449, BSI0581, BSI0639) bound to Protein G magnetic beads with an average size range of 50 nm (MACS® Technology, Miltenyi, Germany). The proteins were then digested off the beads with trypsin in a separation column.

3.6.2. Mass Spectrometry Measurement

The trypsin-digested aliquots were introduced into an LC-MS/MS system. An RP-HPLC (Waters, Milford, MA, USA) device was connected to an Orbitrap-Fusion Lumos (Thermo Fisher Scientific, Waltham, MA, USA) mass spectrometer. For data collection, the top 20 most intense multiply charged precursor ions from the MS1 scan were fragmented in the MS2 scan using Higher-energy C-trap Dissociation (HCD) technology.

3.7. Data Analysis

The raw data were converted to peak list type data using Proteome Discoverer (v 1.4) software (Thermo Scientific, Waltham, MA, USA), and this list was compared to the SwissProt database (version 2017.09.19 containing 555,426 proteins). Protein Prospector (v.5.15.1) software (Thermo Scientific, Waltham, MA, USA) was used for the comparison with the following parameters: enzyme: trypsin with a maximum of two missed cleavages; mass accuracy: 5 ppm for precursor ions and 10 ppm for fragment ions; fixed modifications: cysteine carbamidomethylation; variable modifications: N-terminal acetylation, methionine oxidation, N-terminal glutamine cyclization, serine/threonine/tyrosine phosphorylation, with a maximum of two variable modifications allowed per peptide. The acceptance criteria were an E-value of 0.01 for protein identification and 0.05 for peptide identification. N-glycosylation was searched in the data using the NXT|S, X≠P motifs. N-glycosylation at position 415 was normalized to a strong, unmodified, fully digested peptide present in all samples (LSPIYNLVPVK). (Note: Mass spectrometry experiments were conducted at the Biological Research Centre, Szeged.)

3.8. Antibody Biotinylation

3.8.1. Production of Biotinylated Antibody

To 1 mg of antibody, 26.6 µl of 10 mM EZ-link Sulfo-NHS-LC-Biotin (Thermo Scientific, Waltham, MA, USA) was added. The mixture was incubated for 30 minutes at room temperature with continuous mixing. The biotinylated antibody was purified using Zeba™ Spin Desalting Columns (Thermo Scientific, Waltham, MA, USA). The tip of the column was broken off, placed into a 15 ml centrifuge tube, and centrifuged for 2 minutes at 1000 g and 20°C. The column was then transferred to a clean 15 ml tube, the antibody sample to be biotinylated and 100 µl of ultrapure water were applied to it. The column was centrifuged again for 2 minutes at 1000 g and 20°C. The flow-through contained the biotinylated antibodies.

3.8.2. Validation of Biotinylated Antibody

The degree of biotinylation was determined using the HABA - Pierce Biotin Quantitation Kit (Thermo Scientific, Waltham, MA, USA). The HABA premix stored at 4°C was allowed to warm to room temperature, then dissolved in 100 µl of ultrapure water. The dissolved HABA was applied to a 96-well ELISA plate (160 µl PBS + 20 µl HABA per well). The absorbance (OD) values of the wells were read using a FLUOstar OPTIMA (BMG LabTech, Durham, NC, USA) ELISA plate reader. Then, 20 µl of biotinylated sample was added to the HABA-containing wells. After mixing, the OD values were read again using the plate reader. The Thermo Scientific HABA Calculator program (https://www.thermofisher.com/hu/en/home/life-science/protein-biology/protein-labeling-crosslinking/protein-labeling/biotinylation/biotin-quantitation-kits/haba-calculator.html) was used for the evaluation. The concentration of biotinylated antibodies was measured using a BCA Assay Kit (Thermo Scientific, Waltham, MA, USA). For the BCA measurement, a pre-prepared calibration series and 5 µl of each sample were applied to a 96-well plate, along with 5 µl of a 1 mg/ml IgG standard. Both the samples and the standard were diluted 5-fold with PBS. The A and B components of the kit were mixed in a 50:1 ratio, and 200 µl of the mixture was added to the plate, then incubated for 15 minutes at 37°C. The absorbance values of the plate were obtained using the FLUOstar OPTIMA ELISA reader (BMG LabTech, Durham, NC, USA).

3.9. C9 Epitope Test

The independence of C9 epitopes was tested in a competitive assay on a 96-well ELISA (Costar3690, Corning, NY, USA) plate. Each well of the 96-well plate was coated with 2 µg of C9 protein (Sigma, St. Louis, MO, USA) in 30 µl of carbonate coating buffer (15 mM Na_2CO_3 [Riedel-deHaren, Seelze, Germany], 35 mM $NaHCO_3$ [Riedel-deHaren, Seelze, Germany], pH 9.6) and incubated for one hour at 37°C. The plate was then washed twice with PBS-Tween solution. The next step involved blocking the plate with 60 µl of BSA-PBS-Tween (BSA: Sigma, St. Louis, MO, USA) solution for 30 minutes at 37°C. The plate was washed three times with PBS-Tween solution. Various non-biotinylated, diluted C9 antibodies (BSI0449, BSI0581, BSI0639) in blocking buffer were added to the plate at concentrations of 1 and 10 µg/ml and incubated for 1 hour at 37°C. The plate was then washed twice with PBS-Tween solution, followed by the addition of various C9 antibodies at 1 µg/ml concentration in 30 µl of blocking buffer, and incubated for 1 hour at 37°C. The plate was washed three times, and 5000x diluted

Streptavidin-HRP (Southern-Biotech, Birmingham, AL, USA) in blocking buffer was added for a 30-minute incubation. The plate was washed four times. Then, 30 µl of one-component TMB (Sigma, St. Louis, MO, USA) solution was added to each well, and the reaction was stopped after 1-2 minutes with 4N H_2SO_4 solution. Absorbance was read at 450 nm using the FLUOstar OPTIMA (BMG LabTech, Durham, NC, USA) plate reader.

3.10. Statistical Analysis

Statistical analyses were conducted using the R platform. P-values were calculated with a two-sample t-test ($\alpha = 0.05$) after performing a normality test (Shapiro-Wilk). ROC curves were generated using the pROC R package.

3.11. Visualization

Figures were prepared using the R platform, Excel, IBS Illustrator (version 1.0), and PyMOL (1.7.4.4 Edu) programs.

4. Results

4.1. Selection and Analysis of Anti-C9 Antibodies and Previous Data Analysis

According to the results of our previous measurements, the epitopes of a given protein have different biomarker values in terms of lung cancer association, and these were not uniform in direction, meaning we observed positive, negative, and neutral associations. One of the proteins examined was plasma C9. In the BSI QuantiPlasma library, there are 12 mAbs that recognize C9. From these, we selected three for further examination based on the association of the epitopes recognized by the mAbs. There was a neutral C9 epitope, which did not differentiate between lung cancer and non-lung cancer samples, and there were epitopes that behaved differently in terms of association with plasma samples from lung cancer patients and control individuals, meaning they were positively or negatively associated. We selected one from each of these categories. Statistical analysis showed significant differences between control and lung cancer samples for the BSI0581 and BSI0639 epitopes, while no significant difference was observed for the BSI0449 (neutral) epitope.

4.2. Validation of Anti-C9 Monoclonal Antibodies

4.2.1. Validation on Commercially Available C9

Our experiments were conducted with mAbs produced by BSI; therefore, the first goal of our investigation was to validate the selected BSI antibodies via Western blot following preliminary purity checks. The preselected anti-C9 mAbs (BSI0449, BSI0581, BSI0639) at 2 mg/ml were used at 200x dilution in PBS buffer to recognize commercially available C9 protein (Sigma) as primary antibodies. As secondary antibody, we used GAM-HRP at 2500x dilution. All three anti-C9 proteins recognized the C9 protein in the ~55 kDa molecular weight range, while another negative control antibody specific for albumin (BSI0097) did not recognize C9. The preparation does not contain full-length C9 molecule; the recognized C9 may be a degraded form of the full-length C9, which has a molecular weight of 63.173 kDa.

4.2.2. Examination of C9 in Plasma

After validating the BSI anti-C9 mAbs, we used them to immunoprecipitate pooled whole plasma from control and lung cancer patients. The pools contained samples from 167 control and 207 lung cancer patients. We examined the presence of C9 protein in the immunoprecipitates. The immunoprecipitated control and lung cancer samples were checked via SDS-PAGE followed by Western blot. For the Western blot, we used a 1:1 mixture of BSI0449 and BSI0639. BSI0581 did not work in previous Western blot experiments using whole plasma. As secondary antibody, we used GAM-HRP at 2500x dilution. The results showed that the selected mAbs recognized C9 in the plasma at ~52 kDa and ~68 kDa. These forms could be the previously detected degraded C9 and the full-length plasma C9 with post-translational modifications.

4.3. Examination of Epitope Overlap on C9

We performed a binding competition ELISA test to determine whether the three specific anti-C9 mAbs recognized different epitopes. In this experiment, a 96-well plate was coated with purified, commercially available C9. The next layer consisted of non-biotinylated BSI anti-C9 antibodies (BSI0449, BSI0581, BSI0639) at different concentrations (1 μg/ml and 10 μg/ml). The second antibody was the biotinylated antibody (BSI0449, BSI0581, BSI0639) at 1 μg/ml concentration. (The biotinylation of antibodies was performed in the lab. The degree of biotinylation was checked using the HABA test, and the antibody concentration was checked using the BCA test.) In the first experiment, the primary, non-biotinylated antibody was BSI0449 at different concentrations. As a negative control, bovine serum albumin coating was used instead of C9. As a positive control, the C9 protein was directly detected with the biotinylated anti-C9 mAb. As a technical control, a non-biotinylated antibody was added to the C9, followed by GAM-HRP binding. This experiment was also conducted for BSI0581 and BSI0639 mAbs. Our results suggest that the examined epitopes are independent of each other, but the BSI0449 mAb slightly affects the availability of the BSI0581 and BSI0639 epitopes.

4.4. Analysis of Epitope-Dependent Heterogeneity on C9

4.4.1. Analysis of Epitope Heterogeneity via SDS-PAGE

We examined whether there is a correlation between the molecular heterogeneity associated with the epitopes and the biomarker value. First, we precipitated control and lung cancer plasmas with the three anti-C9 mAbs (BSI0449, BSI0581, BSI0639). The immunoprecipitates were applied to reducing SDS-PAGE gel in different volumes (10 µl and 15 µl), and the gels were stained with Coomassie G-250 and silver staining. There was no significant difference observed between the samples immunoprecipitated with BSI0449 and BSI0581 mAbs. In the samples immunoprecipitated with BSI0639 mAb, a faint band in the ~70 kDa range and two distinct bands in the lung cancer samples were observed.

4.4.2. Analysis of Epitope Heterogeneity via Mass Spectrometry

Due to the differences observed in the co-precipitated proteins (not in the molecular weight of C9), as the resolution of SDS-PAGE was insufficient, we further examined the samples using LC-MS/MS technology. The C9 and its associated proteins anchored to beads with anti-C9 mAbs (BSI0449, BSI0581, BSI0639) were digested with trypsin. The trypsin-digested aliquots were analyzed via mass spectrometry at the Proteomics Laboratory of the Biological Research Centre in Szeged. The examination confirmed that the anti-C9 mAbs react with C9, as 3-13% of the digested peptide aliquots were C9, and this gave 30-45% C9 coverage.

4.5. Data Analysis of LC-MS/MS Results

The different behaviors of the epitopes could be due to genetic modifications such as alternative splicing or alternative translation start or stop sites. I examined whether the observed epitope differences are due to these structural genetic changes in the C9 coding capacity by mapping the MS peptides obtained with different mAbs to the C9 structure found in the Ensembl database (ENST00000263408.5). I searched for C9 post-translational modifications in publications and the UniProt database. During the search, seven different positions (48, 51, 215, 256, 277, 394, 415) showed N-glycosylation, two positions showed C-mannosylation (27, 30), and four positions showed O-glycosylation (24, 26, 32, 258). Signal peptides were missing from our samples in all cases since the C9 was circulating, so secreted C9 in plasma. The distribution

of MS peptides was almost identical. We also compared the number of peptides obtained in different control and lung cancer samples. We collected all C9 peptides from PeptideAtlas (478 peptides) [Human version 2022-01], then arranged these peptides according to their starting positions. Each peptide received an index number in this arrangement. We aligned the C9 MS peptides from our samples to the indexed proteins from PeptideAtlas. After indexing, there was no significant difference observed in the comparison of C9 MS peptides between different samples. The variability of epitopes could also be due to post-translational modifications, so we searched for different glycosylation patterns on C9 MS peptides. During our investigation, we found N-glycosylation at position 415 in all samples. By normalizing the found glycosylation site to a non-modified "strong" peptide, we found differences in the glycosylation ratio between control and lung cancer samples. We also mapped the obtained peptides back to various three-dimensional C9 structures. After mapping to structural images, it was clear that we did not find proteins on the surface that could be a binding point to another protein.

4.6. C9-Associated Plasma Proteins

Protein complexes can dynamically change under different physiological conditions. These changes in protein-protein interactions can mask or expose epitopes. To investigate this, we analyzed the proteins associated with C9. A total of 78 different proteins co-precipitated with C9 in different compositions during the experiments. We selected four of these proteins, which showed at least two peptide differences between control and lung cancer samples when examined with different anti-C9 mAbs. The plasma C9-associated proteins associated to varying degrees. The BSI0449 control and LC samples were strongly associated with C4A, while no C4A peptide was found in the BSI0639 LC sample. Hornerin was not found in the BSI0581 and BSI0639 LC samples. HSP90AB1 peptides were not found in the BSI0449 samples, while they were more strongly associated with the BSI0639 samples. HSPA8 peptides were only found in the BSI0639 samples.

5. Discussion

Epitomics is a field of study first described by our research group, which examines the dynamic epitope changes of proteins in healthy and various pathological conditions. With extensive and global PEP technology, the number of analyzable epitopes, their biomarker value, and functional relevance can be broadly expanded. My work represents the first step in this direction. The plasma epitomic biomarkers discovered so far could potentially be used for diagnostic purposes or patient monitoring. They will not pose additional radiation exposure to patients, and their implementation does not require significant machinery or financial investment. Based on the clinical screening studies conducted with LDCT to date, it is evident that the high number of false-positive results needs to be reduced. Algorithms can be developed using various risk factors, artificial intelligence can be deployed, and all this can be extended to biomarkers and multivariate index approaches that include them.

The C9 protein we studied is the final element of the complement cascade in the innate immune system and part of the terminal complex. In the terminal complex, it polymerizes and joins the previously formed C5b-C6-C7-C8 complex. C9 circulates in plasma in a monomeric form. Clusterin and vitronectin protect it from polymerization in circulation.

C9 has previously been identified as a biomarker in gastric, lung, and rectal tumors. In these publications, only the quantitative presence of the ‘total’ C9 protein was examined.

Based on our research group's previous studies, we selected three anti-C9 antibodies from the QuantiPlasma mAb library that associate differently with lung cancer plasma samples according to sbCIA analysis. Among the profiled epitopes, some were neutral, some positively or negatively associated with lung cancer plasma samples during the experiments.

In our study, we investigated the previously unpublished epitope-dependent molecular heterogeneity of C9 in control and lung cancer plasma samples. We tried to understand the reasons behind the previously discovered phenomenon of different associations. First, we validated the anti-C9 mAbs with commercially available C9 protein. During this process, Western blot analysis confirmed that all three selected anti-C9 mAbs recognize the C9 protein. Although the obtained molecular weight was lower than the literature value, we assume that the pure C9 underwent degradation either during production or storage. After validating our antibodies, we examined the independence of the epitopes using a competitive ELISA method. The three epitopes examined were nearly independent of each other, but the binding of the BSI449 epitope antibody slightly affected the availability and/or structure of the other two epitopes.

To study the C9 protein, we first immunoprecipitated the plasma C9 through the different antibodies. The obtained immunoprecipitate was run on SDS-PAGE and stained with Coomassie Blue and then with silver staining. Based on the results, we determined that the obtained pattern was almost identical. Perhaps the BSI0639 showed more bands, likely because it precipitated less C9, allowing the bands to separate more distinctly. This is also supported by the fact that the C9 band was fainter in the immunoprecipitated samples when examined by Western blot in the case of BSI0639. We also investigated C9 heterogeneity using mass spectrometry. The obtained peptides were mapped back to the C9 protein. Our results showed almost identical peptide distribution in all cases. From this, we concluded that the structure of C9 is likely identical in the samples. Since the structure of C9 is most likely the same in the samples, this cannot explain the initial observation of different epitope associations in control and lung cancer samples.

We continued the experiments in two directions. Using mass spectrometry, we examined the C9 molecular models built with the precipitated peptides, the N-glycosylation at position 415, and the co-precipitated proteins for each epitope and both sample types. Numerous post-translational modifications of C9 have been found in publications and the Uniprot database. Of these modifications, we only found the N-glycosylation at position 415, which was present in all samples. Normalizing the peptide containing this position to an unmodified C9 peptide present in all samples, we found that although the N-glycosylated peptide at position 415 is present in all samples, it is present in different proportions. In the BSI0449 samples, the ratio was almost identical in control and lung cancer samples, while in the BSI0581 and BSI0639, the N-glycosylation ratio was nearly five times higher in favor of the control samples. Examining the co-precipitated proteins, we determined that their ratios varied by antibody and in the control-lung cancer comparison. Based on these findings, we conclude that we identified different proteoforms of C9 with the selected anti-C9 antibodies. Although association with other proteins is not a distinguishing factor among proteoforms, for simplicity, we discuss them together in the following table. We hypothesize that forms 2a and 2b represent variants that arise in the cell under stress or malignant transformation. C9 protein can be produced in various locations, which may also contribute to its different proteoforms.

Our results support the theory that different proteoforms exist. These proteoforms and the complex formation affecting epitope availability may also have biomarker value.

6. Summary

The technology development of our days allows for an increasingly sophisticated understanding of proteins. Describing the proteoforms of proteins and associating them with different states leads to a deeper understanding of protein functions. This knowledge can also be utilized to expand the biomarker function of proteins.

In our previous research, through protein epitope profiling studies, we discovered that protein epitopes can associate in distinct ways with respect to lung cancer. Continuing the previous research of our workgroup, I investigated the epitopes of the C9 protein. We selected three anti-C9 antibodies (BSI0449, BSI0581, BSI0639) from the QuantiPlasma library, with which the epitopes in lung cancer samples showed varying associations (neutral, negative and positive). During my work I investigated the biochemical basis of this phenomenon.

We first validated the anti-C9 antibodies using Western blot. Later, we examined the independence of the three epitopes from each other. The epitopes were largely independent of each other. Using mass spectrometry, we examined the C9 protein from immunoprecipitated samples of control and lung cancerous plasma. We found that the distribution of tryptic peptides was nearly identical, suggesting that the differences are unlikely to originate from the structural difference of C9 variants or proteoforms. Subsequently, we analyzed the proteins co-precipitated of C9. We examined the immunoprecipitated samples using SDS-PAGE with our selected antibodies. The obtained pattern was nearly identical in all samples. We also analyzed the co-precipitated proteins using mass spectrometry. We selected proteins in which there were at least two peptide differences between the control and lung cancer samples. We then identified three proteins that exhibited differences not only among the antibodies but also between the control and lung cancer samples. C4A strongly associated with BSI0449 samples, and it was absent in the BSI0639 lung cancer sample. HSP90AB1 exhibited a strong association with BSI0639 samples and was absent in the BSI0449 samples. HSPA8, on the other hand, was present only in the BSI0639 samples. Investigating the N-glycans; at position 415 of C9, we found N-glycosylation at position 415 in all samples. Although 415-glycosilation was present in all samples, there were differences in the ratio of the 415-N-glycosylation in case of BSI0581 and BSI0639 between control and lung cancer samples. In these cases, the glycosylated peptide was about 5 times more abundant in the control samples. In case of BSI0449, there was no difference between the two states.

Our results suggest that the C9 protein has multiple proteoforms. These proteoforms can be examined through different epitopes. The various proteoforms and the co-precipitated proteins may hold biomarker potential and further biological understanding with respect to lung cancer. My work is the first to provide molecular proof to demonstrate the biological value of epitope-based proteome analysis.

7. New Findings

1. C9 proteins precipitated through different epitopes show differences in lung cancer (LC) and control plasma.

2. The distribution of the mapped C9 peptides suggests the same molecular structure among the different C9 forms precipitated through various epitopes in both LC and control plasma.

3. The co-precipitated proteins exhibit qualitative and quantitative differences between control and LC plasma, indicating variations in spatial associations.

4. The N-glycosylation ratio of C9 at position 415 shows significant differences between control and LC plasma.

8. Keywords

Epitomics, complement C9, epitope profiling, epitope heterogenity, proteoforms, biomarker, lung cancer

9. Acknowledgements

I would like to express my gratitude to my supervisor, Prof. Dr. László Takács, for his guidance and support throughout my work over the past years.

I owe thanks to the colleagues at Biosystems Immunolab Zrt. for their assistance in laboratory work and for providing the financial background for my work.

I am grateful to the former head of the Department of Human Genetics, Prof. Dr. Bálint Nagy, and the current head, Prof. Dr. István Balogh, as well as all the staff at the department, for always being available to answer my questions.

I am also thankful to the former head of the Doctoral School of Molecular Cell and Immunobiology, Prof. Dr. József Tőzsér, and the current head, Prof. Dr. István Balogh, for allowing me to conduct my research within the framework of the doctoral school.

I appreciate the assistance of the colleagues at the Biological Research Centre in Szeged with mass spectrometry analyses and interpretation.

I am grateful to the instructors of the Data-Intensive and Open Science Scholarship program for the knowledge I gained from them.

I would like to thank Prof. Dr. Ildikó Horváth for her always helpful attitude.

I am grateful to Dr. Andrea Furka for her continuous support and expertise over the years.

I appreciate the professional and friendly support of Dr. Gabriella Gellén and (Dr.) Gyula Hoffka.

Last but not least, I would like to thank my mother and my friends for their encouragement, patience, and love.

10. Publications

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/367/2024.PL
Subject: PhD Publication List

Candidate: Ilona Tornyi
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology

List of publications related to the dissertation

1. **Tornyi, I.**, Lázár, J., Pettkó-Szandtner, A., Hunyadi-Gulyás, É., Takács, L.: Epitomics: Analysis of Plasma C9 Epitope Heterogeneity in the Plasma of Lung Cancer Patients and Control Subjects.
   Int. J. Mol. Sci. 24 (18), 1-13, 2023.
   DOI: http://dx.doi.org/10.3390/ijms241814359
   IF: 5.6 (2022)
2. Lázár, J., Kovács, A. L., **Tornyi, I.**, Takács, L., Kurucz, I.: Detection of leucine-rich alpha-2-glycoprotein 1-containing immunocomplexes in the plasma of lung cancer patients with epitope-specific mAbs.
   CBM. 34 (1), 113-122, 2022.
   DOI: http://dx.doi.org/10.3233/CBM-210164
   IF: 3.1

List of other publications

3. Kostyál, L., **Tornyi, I.**, Furka, A.: A szarkopénia mérése komputertomográfiával és jelentősége onkológiai betegeknél.
   Klin. Onkol. 11 (1), 82-86, 2024.
4. Sárközi, A., **Tornyi, I.**, Békési, E., Horváth, I.: Co-Morbidity Clusters in Post-COVID-19 Syndrome.
   J Clin Med. 13 (5), 1-14, 2024.
   DOI: http://dx.doi.org/10.3390/jcm13051457
   IF: 3.9 (2022)
5. **Tornyi, I.**, Horváth, I.: Role of Complement Components in Asthma: a Systematic Review.
   J Clin Med. 13 (11), 1-14, 2024.
   DOI: http://dx.doi.org/10.3390/jcm13113044
   IF: 3.9 (2022)

6. Lieber, A., Makai, A., Orosz, Z., Kardos, T., Susil, J. I., **Tornyi, I.**, Bittner, N.: The role of immunotherapy in early stage and metastatic NSCLC.
Pathol. Oncol. Res. "Accepted by Publisher", 2024.
IF: 2.8 (2022)

7. **Tornyi, I.**, Árkosy, P., Horváth, I., Furka, A.: A new perspective on the proper timing of radiotherapy during CDK4/6 inhibitor therapy in patients with "bone-only" metastatic breast cancer.
Pathol. Oncol. Res. 29, 1-8, 2023.
DOI: http://dx.doi.org/10.3389/pore.2023.1611369
IF: 2.8 (2022)

8. Kostyál, L., **Tornyi, I.**, Sebők, G., Furka, A.: A szarkopénia mérésének radiológiai lehetőségei: Növelhető az onkoterápia hatásossága.
Med. Tribune. 9, 15-16, 2023.

9. Csizmarik, A., Nagy, N., Keresztes, D., Váradi, M., Bracht, T., Sitek, B., Witzke, K., Puhr, M., **Tornyi, I.**, Lázár, J., Takács, L., Kramer, G., Sevcenko, S., Maj-Hes, A., Hadaschik, B., Nyirády, P., Szarvas, T.: Comparative proteome and serum analysis identified FSCN1 as a marker of abiraterone resistance in castration-resistant prostate cancer.
Prostate Cancer Prostatic Dis. [Epub ahead of print], 2023.
DOI: http://dx.doi.org/10.1038/s41391-023-00713-y
IF: 4.8 (2022)

10. Lázár, J., Antal-Szalmás, P., Kurucz, I., Ferenczi, A., Józsi, M., **Tornyi, I.**, Müller, M., Fekete, J. T., Lamont, J., FitzGerald, P., Gall-Debreceni, A., Kádas, J., Vida, A., Tardieu, N., Kieffer, Y., Jullien, A., Guergova-Kuras, M., Hempel, W., Kovács, A. L., Kardos, T., Bittner, N., Csánky, E., Szilasi, M., Losonczy, G., Szondy, K., Gálffy, G., Csada, E., Szalontai, K., Somfay, A., Malka, D., Cottu, P., Bogos, K., Takács, L.: Large-scale plasma proteome epitome profiling is an efficient tool for the discovery of cancer biomarkers.
Mol. Cell. Proteomics. 22 (7), 1-18, 2023.
DOI: http://dx.doi.org/10.1016/j.mcpro.2023.100580
IF: 7 (2022)

11. **Tornyi, I.**, Furka, A.: Új biomarkerek a tüdőrákok diagnosztikájában: Lehetőség a tüdőrákok korai felismeréséhez.
Med. Tribune. 9, 9-10, 2023.

12. Keresztes, D., Csizmarik, A., Nagy, N., Módos, O., Fazekas, T., Bracht, T., Sitek, B., Witzke, K., Puhr, M., Sevcenko, S., Kramer, G., Shariat, S., Küronya, Z., Takács, L., **Tornyi, I.**, Lázár, J., Hadaschik, B., Lászik, A., Szűcs, M., Nyirády, P., Szarvas, T.: Comparative proteome analysis identified CD44 as a possible serum marker for docetaxel resistance in castration-resistant prostate cancer.
J. Cell. Mol. Med. 26 (4), 1332-1337, 2022.
DOI: http://dx.doi.org/10.1111/jcmm.17141
IF: 5.3

13. Furka, A., Nagy, Z., Szabó, I., Fekete, G., Kelemen, Á., Bolobás, G., Sebők, G., Molnár, T., Árvai, J., **Tornyi, I.**, Kostyál, L., Révész, J., Hauser, P.: Full Body Surface Coverage with Water-Equivalent Bolus as Novel Technique for Total Body Irradiation before Hematopoietic Stem Cell Transplantation in Pediatric Acute Lymphoid Leukemia.
Children. 9 (11), 1-10, 2022.
DOI: http://dx.doi.org/10.3390/children9111740
IF: 2.4

14. Csizmarik, A., Keresztes, D., Nagy, N., Bracht, T., Sitek, B., Witzke, K., Puhr, M., **Tornyi, I.**, Lázár, J., Takács, L., Kramer, G., Sevcenko, S., Maj-Hes, A., Jurányi, Z., Hadaschik, B., Nyirády, P., Szarvas, T.: Proteome profiling of enzalutamide-resistant cell lines and serum analysis identified as marker of resistance in castration-resistant prostate cancer.
Int. J. Cancer. 151 (8), 1405-1419, 2022.
DOI: http://dx.doi.org/10.1002/ijc.34159
IF: 6.4

15. Furka, A., Simkó, C., Kostyál, L., Szabó, I., Valikovics, A., Fekete, G., **Tornyi, I.**, Oross, E., Révész, J.: Treatment Algorithm for Cancerous Wounds: a Systematic Review.
Cancers (Basel). 14 (5), 1-12, 2022.
DOI: http://dx.doi.org/10.3390/cancers14051203
IF: 5.2

Total IF of journals (all publications): 53,2
Total IF of journals (publications related to the dissertation): 8,7

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

17 June, 2024

DEBRECENI EGYETEM
Egyetemi és Nemzeti Könyvtár