



## **Discriminating Benign from Malignant Lung Diseases Using Plasma Glycosaminoglycans and Cell-Free DNA**

Downloaded from: <https://research.chalmers.se>, 2024-11-05 12:15 UTC

Citation for the original published paper (version of record):

Qvick, A., Bratulic, S., Carlsson, J. et al (2024). Discriminating Benign from Malignant Lung Diseases Using Plasma Glycosaminoglycans and Cell-Free DNA. *International Journal of Molecular Sciences*, 25(18).  
<http://dx.doi.org/10.3390/ijms25189777>

N.B. When citing this work, cite the original published paper.



Article

# Discriminating Benign from Malignant Lung Diseases Using Plasma Glycosaminoglycans and Cell-Free DNA

Alvida Qvick <sup>1,\*</sup>, Sinisa Bratulic <sup>2</sup>, Jessica Carlsson <sup>3</sup>, Bianca Stenmark <sup>1</sup>, Christina Karlsson <sup>4</sup>,  
Jens Nielsen <sup>2,5</sup>, Francesco Gatto <sup>2,6</sup> and Gisela Helenius <sup>1</sup>

<sup>1</sup> Department of Laboratory Medicine, Faculty of Medicine and Health, Örebro University, 701 82 Örebro, Sweden

<sup>2</sup> Department of Life Sciences, Chalmers University of Technology, 412 96 Gothenburg, Sweden

<sup>3</sup> Department of Urology, Faculty of Medicine and Health, Örebro University, 701 82 Örebro, Sweden

<sup>4</sup> School of Health Sciences, Örebro University, 701 82 Örebro, Sweden; christina.karlsson@oru.se

<sup>5</sup> BioInnovation Institute, Ole Maaløes Vej 3, 2200 Copenhagen, Denmark

<sup>6</sup> Department of Oncology-Pathology, Karolinska Institute, 171 77 Stockholm, Sweden

\* Correspondence: alvida.qvick@regionorebrolan.se

**Abstract:** We aimed to investigate the use of free glycosaminoglycan profiles (GAGomes) and cfDNA in plasma to differentiate between lung cancer and benign lung disease, in a cohort of 113 patients initially suspected of lung cancer. GAGomes were analyzed in all samples using the MIRAM<sup>®</sup> Free Glycosaminoglycan Kit with ultra-high-performance liquid chromatography and electrospray ionization triple quadrupole mass spectrometry. In a subset of samples, cfDNA concentration and NGS-data was available. We detected two GAGome features, 0S chondroitin sulfate (CS), and 4S CS, with cancer-specific changes. Based on the observed GAGome changes, we devised a model to predict lung cancer. The model, named the GAGome score, could detect lung cancer with 41.2% sensitivity (95% CI: 9.2–54.2%) at 96.4% specificity (95% CI: 95.2–100.0%, n = 113). When we combined the GAGome score with a cfDNA-based model, the sensitivity increased from 42.6% (95% CI: 31.7–60.6%, cfDNA alone) to 70.5% (95% CI: 57.4–81.5%) at 95% specificity (95% CI: 75.1–100%, n = 74). Notably, the combined GAGome and cfDNA testing improved the sensitivity, compared to cfDNA alone, especially in ASCL stage I (55.6% vs 11.1%). Our findings show that plasma GAGome profiles can enhance cfDNA testing performance, highlighting the applicability of a multiomics approach in lung cancer diagnostics.

**Keywords:** cfDNA; GAGome; glycosaminoglycans; lung cancer; multiomics



**Citation:** Qvick, A.; Bratulic, S.; Carlsson, J.; Stenmark, B.; Karlsson, C.; Nielsen, J.; Gatto, F.; Helenius, G. Discriminating Benign from Malignant Lung Diseases Using Plasma Glycosaminoglycans and Cell-Free DNA. *Int. J. Mol. Sci.* **2024**, *25*, 9777. <https://doi.org/10.3390/ijms25189777>

Academic Editor: Mariaenrica Tinè

Received: 30 July 2024

Revised: 5 September 2024

Accepted: 8 September 2024

Published: 10 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The use of high-throughput sequencing to analyze genomes and their aberrations has advanced diagnostics, predictive testing, and monitoring for many types of cancer. To avoid the challenges of obtaining tissue biopsies for cancers like lung cancer (LC), liquid biopsy has emerged as a promising alternative. Numerous articles have been published on the sequencing of circulating cell-free DNA (cfDNA) for such cancers [1]. Despite technical advances, which have improved the sensitivity of genomics-only liquid biopsy assays, many cancers remain undetected. This can be due to technical as well as biological reasons, since the amount of circulating tumor DNA released is very heterogeneous between different tumors [2]. To overcome this limitation, the field has moved beyond genomics to a combination of different omics techniques, referred to as multiomics, including combined information from genomics, transcriptomics, proteomics, methylomics, metabolomics, extracellular vesicles, and circulating tumor cells, among others [3–8].

Glycosaminoglycans (GAGs) are unbranched linear polysaccharides that can be divided into four main classes: chondroitin sulfate, heparan sulfate, keratan sulfate, and hyaluronic acid [9]. The sulfation and glycosylation patterns in GAG chains can vary

widely, leading to very high structural and functional diversity in the resulting GAG molecules [10]. GAGs play crucial roles in cellular functions, including maintaining the extracellular matrix structure and providing hydration to cells. They are also integral to the immune response [11,12] and tissue homeostasis, as well as cell growth, proliferation, differentiation, and adhesion [13]. Notably, GAGs have been implicated in different aspects of cancer development and progression due to their involvement in the tumor microenvironment through their interactions with growth factors, growth factor receptors, and cytokines [14]. The disaccharide compositions of GAGs in tumor tissue, plasma, and urine have been shown to be altered in several cancer types such as breast, prostate, gastric, and renal cell carcinoma [15–19]. Conversely, only a few studies have measured the structural profiles of GAG disaccharides (or GAGome) in tissue or liquid biopsies in LC [20,21], indicating the need for further research.

This study aimed to explore the plasma GAGome’s potential for differentiating LC from non-malignant lung diseases via liquid biopsy. Additionally, we investigated whether combining GAGome and cfDNA data could increase the sensitivity and specificity of LC detection.

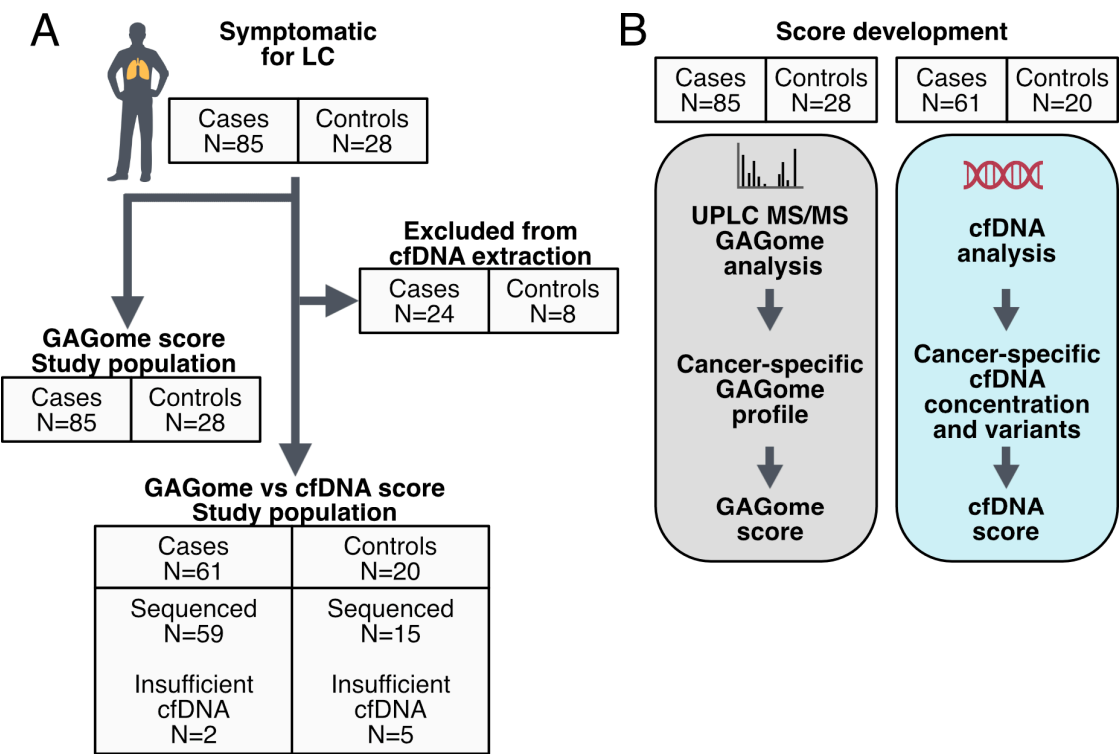
## 2. Results

### 2.1. Cohort Characteristics

The cohort comprised patients referred to the clinic under suspicion of LC ( $n = 113$ ). Of these, 85 cases were subsequently diagnosed with LC and were further stratified into non-small cell LC (NSCLC,  $n = 77$ ) or small cell LC (SCLC,  $n = 8$ ). Patients diagnosed with a benign lung disease ( $n = 28$ ) served as controls (Table 1, Figure 1). Cases were slightly older than controls (mean 70.4 vs. 67.0 years) and more evenly distributed among the sexes (52.9% females vs. 60.7% females), with neither characteristic showing a significant difference. A history of smoking was more prevalent among cases compared to controls (85.9% vs. 60.8%,  $p = 0.047$ ). The predominant tumor stage was stage IV (61.2%), with a mean tumor size of 4.48 cm.

**Table 1.** The clinical characteristics of the cohort. Cases include patients diagnosed with lung cancer, irrespective of histology, and controls include patients diagnosed with different benign diseases of the lung. NA: not applicable.

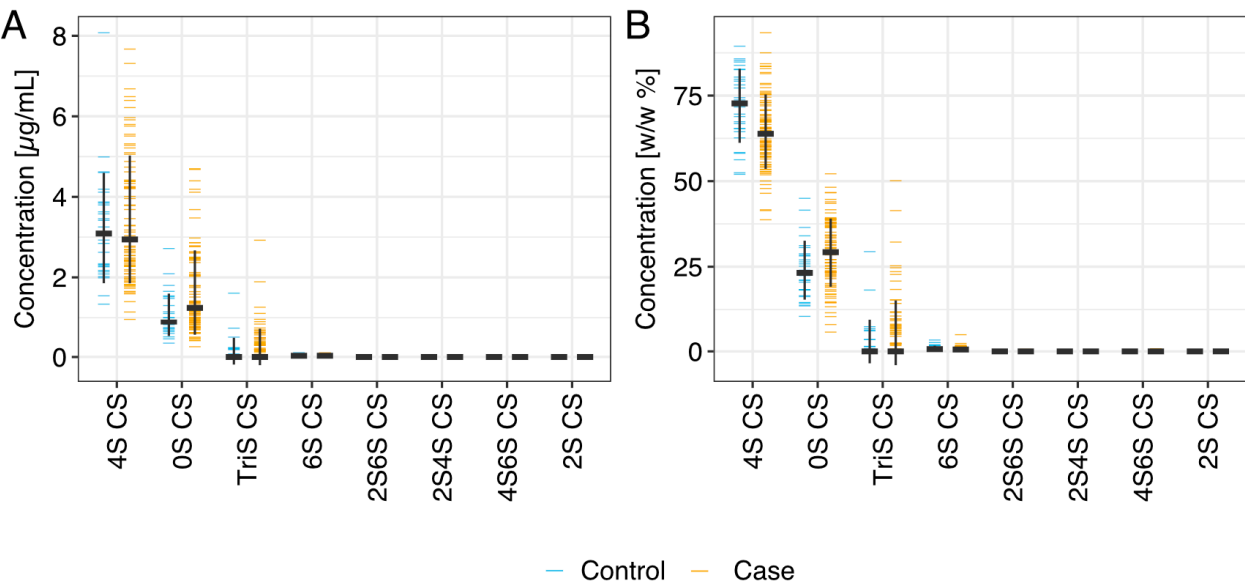
	Case (n = 85)	Control (n = 28)	Case vs. control, <i>p</i> -Value	Overall (n = 113)
<b>Age</b>				
Mean (SD)	70.4 (8.5)	67.0 (14.4)	0.246	69.6 (10.3)
<b>Gender</b>				
Female	45 (52.9%)	17 (60.7%)	0.473	63 (55.3%)
Male	40 (47.1%)	11 (39.3%)		51 (44.7%)
<b>Smoking</b>				
Smoker	30 (35.3%)	5 (17.9%)	0.047	35 (31.0%)
Ex-Smoker	43 (50.6%)	12 (42.9%)		55 (48.7%)
Never-Smoker	12 (14.1%)	9 (32.1%)		21 (18.6%)
Missing	0 (0%)	2 (7.1%)		2 (1.8%)
<b>Tumor size (cm)</b>				
Mean (SD)	4.5 (2.4)	NA (NA)		4.5 (2.4)
Median [Min, Max]	4.0 [0.9, 12.5]	NA [NA, NA]		4.0 [0.9, 12.5]
Missing	4 (4.7%)	28 (100%)		32 (28.3%)
<b>IASLC stage</b>				
I	9 (10.6%)	0 (0%)		9 (8.0%)
II	6 (7.1%)	0 (0%)		6 (5.3%)
III	18 (21.2%)	0 (0%)		18 (15.9%)
IV	52 (61.2%)	0 (0%)		52 (46.0%)



**Figure 1.** Study overview. (A) Cohort inclusion for GAGome and cfDNA analysis. (B) Development of scores using GAGome and cfDNA.

2.2. Correlation between Plasma GAGomes and Lung Cancer Diagnosis

The plasma GAGome, which comprises the structural characterization of 17 different GAG disaccharides, was analyzed successfully in all samples (n = 113). 4S CS and 0S CS were the only GAG disaccharides with a median detected concentration above 0.1 µg/mL. Therefore, we focused the subsequent analysis on these two independently measured features and the corresponding four derived features (Section 4 Materials and Methods, Table S2, Figure 2).



**Figure 2.** Plasma GAGome profiles. Concentrations (A) and fractions (B) of plasma GAGs disaccharides in cases (n = 85) and controls (n = 28). CS: chondroitin sulfate.

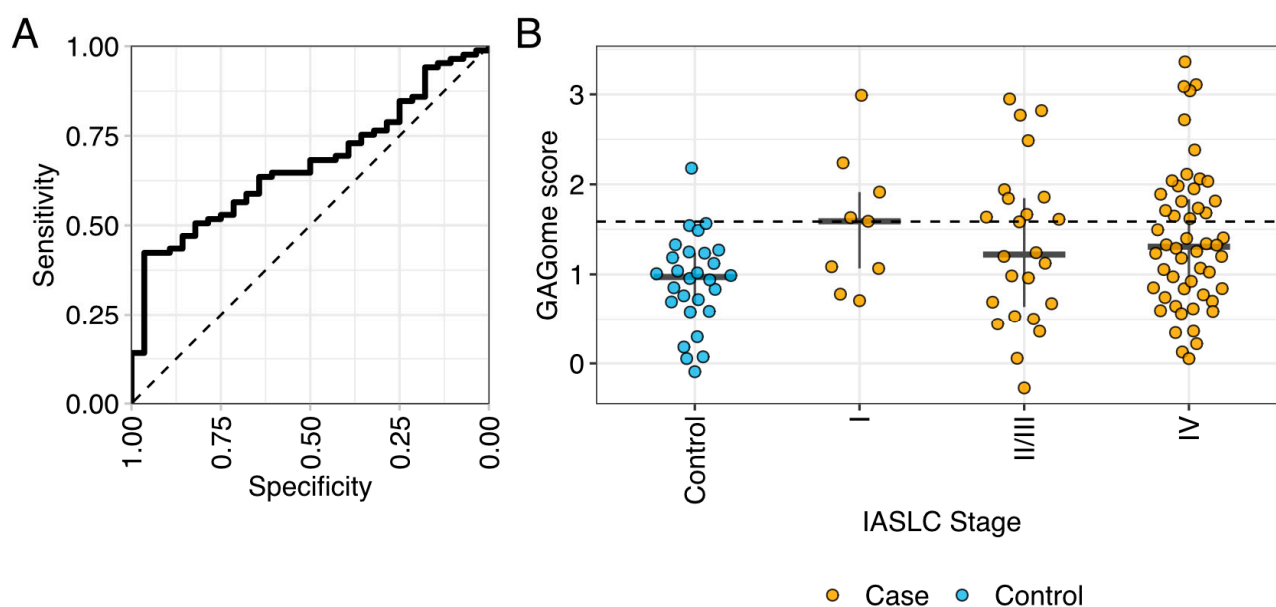
Compared to the controls, the cases with cancer had a nominally higher median concentration of 0S CS, and a lower concentration of 4S CS (Figure 2, Table S2). Concomitantly, we found that the composition of CS in cases with cancer shifted towards a lower 4S CS fraction and a higher 0S CS fraction (Figure 2).

Using Bayesian linear regression and equivalence testing, we determined that two plasma GAGome features were credibly associated with LC after adjustment for batch effects and sample age (Figure S1). Specifically, an increase in the plasma concentration of 0S CS and a reduction in the fraction of 4S CS were credibly correlated with LC after adjustment for confounders.

### 2.3. GAGome and cfDNA Score

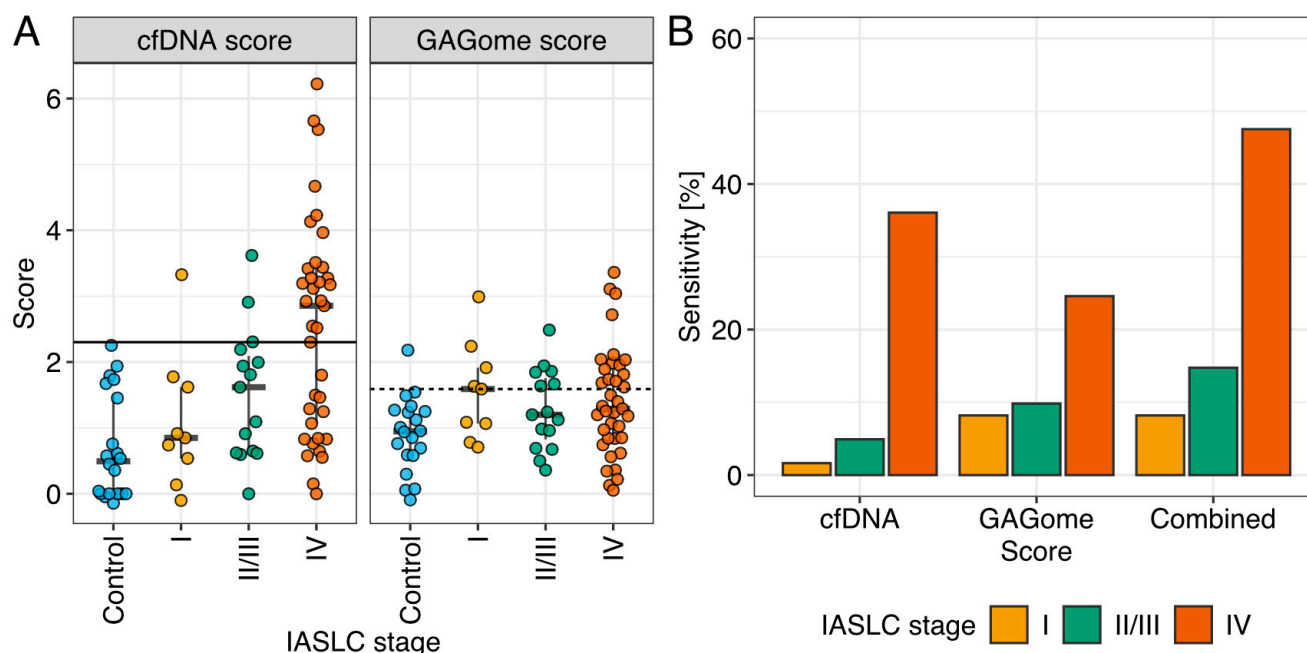
We next sought to develop a plasma GAGome LC score (or simply, GAGome score) for discriminating between LC cases and controls. We fitted a Bayesian logistic model to predict cancer (Figure S2, Supplementary Materials), where the model's output, referred to as the GAGome score, corresponded to the log-odds ratio of having LC.

The GAGome score had a 41.2% sensitivity (95% CI: 9.2–54.2%) to LC at a 96.4% specificity (95% CI: 95.2–100%) with an AUC of 0.67 (95% CI: 0.56–0.77, Figure 3, Table S3).



**Figure 3.** The performance of the GAGome score in discriminating between cases and controls in (A) the whole cohort and (B) by IASLC stage. Controls are shown in blue ( $N_{\text{Control}} = 28$ ) and cases in orange ( $N_{\text{Case}} = 85$ ;  $N_{\text{Stage I}} = 9$ ,  $N_{\text{Stage II/III}} = 24$ ,  $N_{\text{Stage IV}} = 58$ ). IASLC: International Association for the Study of Lung Cancer.

Next, cfDNA data was analyzed. Neither cfDNA concentration nor the number of cfDNA variants were correlated with the GAGome features included in the GAGome score (0S CS concentration and 4S CS fraction, Figure S3). A cfDNA score was generated in the subset samples for which both cfDNA concentration and the number of variants were available ( $n = 74$ ). Specifically, we fitted a Bayesian logistic model to predict case vs. control, using cfDNA concentration and the number of cfDNA variants as predictors. The model's output, referred to as the cfDNA score, corresponded to the log-odds ratio of having LC. We treated samples where cfDNA concentration was available but insufficient for cfDNA variant analysis ( $n = 7$ ) as cfDNA score negative when evaluating the cfDNA score performance. The cfDNA score had an AUC of 0.80 (95% CI: 0.685–0.903) and 42.6% sensitivity (95% CI: 31.7–60.6%) at 100% specificity (Figure 4A).



**Figure 4.** The performance of the combined test. (A) The performance and cut-offs of the cfDNA and GAGome scores separately, and (B) the sensitivity when combined (by IASLC stage). Color scheme is: controls (blue), stage I (yellow), stage II/III (green), and stage IV (red). IASLC: International Association for the Study of Lung Cancer.

We envisioned a diagnostic pathway that relies on sequential cfDNA and GAGome measurements in plasma, specifically using the GAGome score for potential cfDNA score false negatives. This combined test would be positive if either the cfDNA or the GAGome score was positive. We allowed for one false positive and set the threshold for cfDNA positivity at 100% specificity and GAGome score positivity at 95% specificity (Figure 4A). The sensitivity of the combined test increased to 70.5% (95% CI: 57.4–81.5%) at a specificity of 95% (95% CI: 75.1–100%, Figure 4B). Notably, in this diagnostic pathway with 95% specificity overall (one false positive), the GAGome score contributed to a higher sensitivity towards stage I (55.6% vs. 11.1% for cfDNA alone, Table S4) and stage II–III (40% vs. 20% for cfDNA alone, Table S4), and could correctly reclassify 17 out of 19 cases that were falsely marked as negative when using the cfDNA score alone (Table S5).

### 3. Discussion

In this study, we aimed to use GAGomes to discriminate between LC and benign lung diseases and to explore whether the addition of cfDNA data could further increase the sensitivity of the developed score. We used the detected features, 0S CS and 4S CS, to build a GAGome score that reached 41.2% sensitivity at 96.4% specificity. We next envisioned a test that combines the GAGome score with cfDNA data to increase the diagnostic potential for LC. In the subset of patients with cfDNA data available, the combined test could diagnose LC with 70.5% sensitivity at 95% specificity. To the best of our knowledge, this is the first study to develop a multiomics test for cancer combining GAGs and cfDNA.

Research on GAGs in LC has been limited, with most studies focusing on tissue-based analysis. An increase in total CS has been observed in tumor tissue compared to normal lung tissue [20–22]. However, the results regarding CS sulfation patterns have been inconsistent. Pál et al. and Balbisi et al. reported lower 0S CS and higher 4S CS in tumors, while Li et al. found the opposite for 4S CS. Discrepancies between tissue and plasma can be explained by the cell origin of the molecules analyzed. Mattox et al. investigated the contribution of different cells to cfDNA in plasma and found that, even in cancer patients, over 70% of cfDNA originated from leukocytes, and only 2.2% of the fragments



in LC patients originated from the lung [23]. They concluded that these results reflect the systemic effect of the tumor on the body, suggesting that what is detected is not primarily the tumor itself, but the effect it causes. This concept can be applied to other circulating biomarkers, such as GAGs, indicating that plasma tests should be interpreted separately from tissue tests. Circulating GAGs in plasma have only been reported once previously, where an elevated 0S CS concentration and a lower 4S CS fraction were detected in LC cases compared to controls [24].

We found that the GAGome alone had a sensitivity of 40% for detecting LC, which was comparable to the 42.6% sensitivity of cfDNA. Given that GAGomes were uncorrelated with cfDNA measurements and that the GAGome score had a similar sensitivity across all IASLC stages, we speculated that the two scores could be used in a combined multiomics test for LC. By combining scores derived from cfDNA and GAGs into a multiomic test, the sensitivity of differentiation between LC and non-malignant lung diseases could increase to 70.5%, highlighting the potential of a multiomics approach.

The effectiveness of using multiomics for cancer detection was notably proposed by Cohen et al. through CancerSEEK, which analyzes mutations in cfDNA and protein levels [3]. They reported an overall sensitivity of 70% across all tumor types and about 59% for LC specifically. Since then, several other studies have explored the potential of multiomics as a diagnostic tool.

For instance, Wang et al. [4] tested a combination of cfDNA variants, proteins, and fragmentomics on a cohort of colorectal, esophageal, gastric, liver, lung, and ovarian cancers [25]. At 98% specificity, cfDNA variants alone showed a sensitivity of 46%. This increased to 60% with the addition of proteins, and further to 66% with fragmentomics. However, for LC alone, the sensitivity of the combined model only reached 38.5%. Chen et al. combined a mutational score, methylation, and serum CEA levels to distinguish between LC and benign lung nodules, achieving 76.9% sensitivity at a modest specificity of 58.3%. At the 95% specificity level used in our study, the sensitivity dropped to around 30% [6]. D'Ambrosi et al. reported a promising 68% sensitivity at 95% specificity for diagnosing LC using platelet-derived circRNA and mRNA [26]. Notably, their control group primarily consisted of asymptomatic individuals, similar to Wang et al. and CancerSEEK [3,25]. Our study differed in design by using patients referred to the clinic with lung-related symptoms qualifying for LC evaluation, rather than selecting a specific control group. This enhances the clinical relevance of our findings, as it better reflects real-world diagnostic challenges. Although our control cohort was more suitable compared to several other multiomics studies, it is important to note that most of these studies had a skew towards the lower stages of LC, which may have negatively impacted their reported sensitivity.

The only detectable plasma GAGome features in this cohort were 0S CS and 4S CS. This is consistent with previous findings using this kit, both in healthy individuals and cancer patients [24,27]. Other studies have detected additional GAGs, such as HA, in plasma [28]. However, the methods used in those studies included proteolytic digestion, while the method used here is degradation-free, detecting only GAGs that circulate freely in the analyzed liquid. Saito et al. tested the affinity of plasmatic proteins to different GAGs and showed that approximately 7.5% of the proteins are bound to HS and dermatan sulfate, while only 0.25% are bound to CS [29]. This suggests that CS has a higher probability of circulating freely in plasma, supporting our findings.

The limitations of our study include the relatively small sample size and lack of external cohorts for validation. A methodological limitation is that blood was collected in tubes optimized for the stabilization of cells and extracellular RNA, and not immediately processed, as the stabilization reagents are effective for several days. However, GAGs are primarily degraded by highly substrate-specific enzymes in the lysosome [30]. Since the tubes' main function is to stabilize cells, lysosomal activity in the plasma fraction of the sample is unlikely. While an unspecific degradation in plasma cannot be ruled out, the cancer-specific alterations to the GAGomes were resistant to adjustment for sample age.

In conclusion, we have shown that free CS can be detected in the plasma of LC patients. From this, we developed a highly specific and sensitive multiomics score by combining these data with cfDNA data from NGS analysis, effectively differentiating between LC and benign lung diseases. We envision this score as a natural companion diagnostic to radiography, as a broad NGS is highly relevant for LC due to the increasing number of targeted therapies available for this condition. Adding a GAGome analysis would be comparatively inexpensive, easily performed on a small aliquot of the same blood sample used for NGS analysis, and could specifically increase sensitivity towards lower stages of cancer without affecting the false positive rate.

#### 4. Materials and Methods

##### 4.1. Study Design and Cohort Characteristics

This study was reported in compliance with the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines [31] (Table S1). Study participants were enrolled into the study during routine clinical investigation at the lung clinic at Örebro University Hospital between February 2016 and February 2019. Patients investigated for suspected LC were included in the study and formed a consecutive series. To ensure the most accurate diagnosis, all LC cases were histologically confirmed using tumor tissue. Patients who were diagnosed with cancer originating somewhere other than the lung, whose samples were not collected prior to the start of treatment, or who had inadequate tumor material for diagnosis were excluded from the study. Participants gave written informed consent before inclusion and the study was approved by the regional ethics committee board in Uppsala (Approval 2015-400, 2021-01478).

Tumors were staged and histologically classified according to the guidelines of the International Association for the Study of Lung Cancer (IASLC) and the World Health Organization nomenclature, respectively [32,33]. Patients with benign lung diseases mainly consisted of those with different pulmonary obstructive diseases, inflammatory conditions in the lung, fibrosis in the lung, or benign lung nodules.

##### 4.2. Plasma Collection and Isolation

Blood was collected in Cell-Free RNATM BCT tubes (Streck, Omaha, NE, USA) and plasma was retrieved by two-step centrifugation at  $2000\times g$  for 10 min followed by  $16,000\times g$  for 10 min. Plasma was stored at  $-80\text{ }^{\circ}\text{C}$  until preparation for analysis. Samples were thawed on ice and an aliquot of each sample was taken for further analysis.

##### 4.3. GAGome Measurements

Plasma GAGome measurements were performed retrospectively in a single-blinded Good Laboratory Practice (GLP)-compliant central laboratory using the MIRAM® Free Glycosaminoglycan Kit (Product No. FRUOV1, Elypta AB, Solna, Sweden), which is a standardized kit for GAG extraction, detection, and quantification by ultra-high-performance liquid chromatography (UHPLC), coupled with an electrospray ionization triple quadrupole mass spectrometry system (ESI-MS/MS, Acquity I-class Plus Xevo TQ-S micro, Waters® Corporation, Milford, MA, USA). A single UHPLC column equipped with a pre-column guard (Waters® ACQUITY UPLC BEH C18 VanGuard Pre-column) was sufficient to analyze all samples in this study, with no quality deterioration observed over time. The analytical performance characteristics of the kit have been previously described [34].

In short, the kit is based on a method by Volpi et al. [35]. The assay consists of the enzymatic depolymerization of GAGs from the sample into disaccharides by Chondroitinase ABC and Heparinase I-II-III. The method omits proteolytic digestion, thereby limiting the derived depolymerized GAGs to the protein-free fraction, or free GAGs. Following depolymerization, disaccharides are labeled using 2-aminoacridone and injected into an UHPLC-MS/MS for separation and detection. The peaks of the 17 disaccharides are acquired using multiple reaction monitoring analysis, implemented in the mass spectrometry



software (Waters® TargetLynx). The chromatographic conditions and MS configuration are set in accordance with the kit's instructions for use.

Each sample was measured singly. The so-measured GAGome consisted of the absolute concentrations of seventeen GAG disaccharides, corresponding to eight different sulfation patterns of chondroitin sulfate (CS) and heparan sulfate (HS), and one hyaluronic acid (HA) disaccharide. Specifically, we quantified eight CS disaccharides (0S CS, 2S CS, 6S CS, 4S CS, 2S6S CS, 2S4S CS, 4S6S CS, and TriS CS) and eight HS disaccharides (0S HS, 2S HS, 6S HS, NS HS, NS6S HS, NS2S HS, 2S6S HS, and TriS HS). We expanded the GAGome to include an additional 22 calculated features informative of GAG biology: (a) the total CS and total HS concentration as the sum of the corresponding disaccharide concentrations; (b) the CS charge [-] and HS charge [-] as the weighted sum of the sulfated disaccharides, where the weight is the count of sulfo groups in each disaccharide; (c) two ratios (4S CS/0S CS and 6S CS/0S CS); and (d) the relative concentration (or mass fraction, in  $\mu\text{g}/\mu\text{g}\%$ ) of each of the sixteen CS and HS disaccharides, found by normalizing each absolute concentration by the total CS and HS concentration, respectively. For each sample, the GAGome consisted maximally of 39 features.

We considered a GAGome feature detectable in plasma if the median concentration across all samples was above  $0.1 \mu\text{g}/\text{mL}$  [34]. GAGome features that did not fulfill this criterion were excluded from downstream analyses.

#### 4.4. cfDNA Data

For 81 of the 113 samples, cfDNA concentration measurement was available. The cfDNA was extracted from 4 mL plasma using the QIAasymphony DSP Circulating DNA kit on the QIAasymphony SP system (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The concentration was measured using a dsDNA HS assay kit (Thermo Fisher, Waltham, MA, USA) on a Qubit 2.0 Fluorometer (Thermo Fisher). Out of the samples with available cfDNA concentration measurements, 74 samples also had NGS data available, which have been published previously [36]. The NGS panel used was the AVENIO ctDNA Surveillance kit (Roche Diagnostics, Rotkreuz, Switzerland), which includes 197 cancer-relevant genes.

#### 4.5. Statistical Analysis

Continuous data were presented using mean and median values, while categorical data were presented using absolute and relative frequencies. Differences between LC patients and controls in terms of general clinical characteristics were investigated using  $\chi^2$ -tests for categorical variables and Student's *t*-tests for continuous data.

##### 4.5.1. GAGome Analysis

We compared the levels of each detectable GAGome feature in cancer versus control participants using Bayesian estimation with practical equivalence testing [37]. First, we fitted a Bayesian linear regression model where each individual GAGome feature was standardized and modelled as a normally distributed response variable and the disease state (case vs. control) was the only binary predictor. Second, we controlled the correlation between GAGome measurements in cancer vs. controls for technical variation by fitting Bayesian linear models as above, but including an experimental batch (binary) and sample age (continuous, in months) as predictors. We used a *t*-distribution centered on 0 with 7 degrees of freedom, and a scale = 2 for all the priors in all the Bayesian models. We fit the models using the rstanarm package (ver 2.21.3) in R (ver. 4.2.1). The convergence and stability of the Bayesian sampling was assessed using R-hat, which should be below 1.01 [38], and Effective Sample Size (ESS), which should be greater than 1000 [39]. The same convergence criteria were used for all Bayesian models.

A GAGome feature was considered credibly associated to case–control status if the 95% confidence interval (CI) of the difference in means did not include zero, and no more than 10% of the CI passed into the Region of Practical Equivalence (ROPE) [37]. The ROPE-

interval was defined as  $[-0.1, 0.1]$  of the standardized mean, corresponding to a negligible effect size [40]. GAGome features that were credibly associated with case–control status after adjusting the linear regression model for experimental batch and sample age were further analyzed as predictors of cancer during the GAGome score development.

#### 4.5.2. GAGome Score Development

A Bayesian logistic model was fitted to predict case vs. control using the absolute concentration of OS CS (in  $\mu\text{g/mL}$ ) and the fraction of 4S CS (in  $\mu\text{g}/\mu\text{g}\%$ ) as predictors. The predictors were batch-normalized before fitting the model. Markov chain Monte Carlo (MCMC) sampling was performed with four chains of 10,000 iterations and a warmup of 5000. The output of the model is referred to as the plasma GAGome LC score, or simply GAGome score.

Model metrics were investigated using a bootstrap analysis with 5000 bootstraps, under a constraint of minimum 95% specificity, which was deemed potentially clinically useful. The final model's performance in detecting case vs. control was evaluated by calculating sensitivity at 95% specificity and the area under the curve (AUC) of the model.

#### 4.5.3. cfDNA Score Development

Bayesian logistic models (estimated using MCMC sampling) were fitted to predict case–control status using cfDNA concentration (logarithmic) in combination with the number of cfDNA variants. The output of the model is referred to as the cfDNA score.

#### 4.5.4. Combined GAGome and cfDNA Test Pathway

A combined test was envisioned as a diagnostic pathway that sequentially uses the cfDNA score and the GAGome score to render a diagnostic decision, i.e., “combined test positive” vs. “combined test negative”. To cumulatively retain the 95% specificity for the combined test, as was rationalized for the GAGome score, the GAGome specificity was kept at 95% while the specificity for the cfDNA score was set to 100%. Specifically, the GAGome score, when positive, could be used to reclassify cfDNA score false negative samples as “combined test positive”. The testing procedure is as follows: (1) calculate the cfDNA score cut-off at which a minimum of 100% specificity is achieved; (2) calculate the GAGome score cut-off at which a minimum of 95% specificity is achieved; (3) for each sample, consider it positive for the cfDNA score if the sample score is above the cut-off—this also applies for the GAGome score; (4) for each sample, consider it positive for the combined test if at least one between the cfDNA and GAGome score is positive; (5) mark samples which are cfDNA score negative but GAGome score positive as “reclassified positive”.

**Supplementary Materials:** The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25189777/s1>.

**Author Contributions:** Conceptualization, A.Q. and G.H.; methodology, A.Q., S.B., F.G., J.C. and G.H.; software, S.B.; formal analysis, A.Q., S.B., J.C., B.S., C.K., F.G., J.N. and G.H.; investigation, A.Q., S.B., and J.C.; resources, A.Q., S.B., F.G., J.N. and G.H.; data curation, A.Q. and S.B.; writing—original draft preparation, A.Q. and S.B.; writing—review and editing, A.Q., S.B., F.G., J.C., B.S., C.K., J.N. and G.H.; visualization, A.Q. and S.B.; supervision, G.H.; project administration, A.Q. and G.H.; funding acquisition, A.Q. and G.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Nyckelfonden-Örebro University Hospital Research Foundation, the Lions Fund for Cancer Research Uppsala-Örebro, and the Uppsala-Örebro Regional Research Council.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the regional ethics committee board in Uppsala (Approval 2015-400, 2021-01478).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data is available from the authors upon reasonable request.

**Conflicts of Interest:** F. Gatto and J. Nielsen are shareholders in Elypta AB. F. Gatto and S. Bratulic are employed by Elypta AB. J. Nielsen is a board member at Elypta AB. Elypta AB has a commercial interest in part of the technology described in this study. The remaining authors have no conflicts to declare.

## References

- Chen, Z.; Miao, H.; Zeng, Q.; Xu, S.; Chen, Z.; Liu, K. Circulating cell-free DNA as a diagnostic and prognostic biomarker for non-small-cell lung cancer: A systematic review and meta-analysis. *Biomark. Med.* **2020**, *14*, 587–597. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bettegowda, C.; Sausen, M.; Leary, R.J.; Kinde, I.; Wang, Y.; Agrawal, N.; Bartlett, B.R.; Wang, H.; Luber, B.; Alani, R.M.; et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci. Transl. Med.* **2014**, *6*, 224ra24. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cohen, J.D.; Li, L.; Wang, Y.; Thoburn, C.; Afsari, B.; Danilova, L.; Douville, C.; Javed, A.A.; Wong, F.; Mattox, A.; et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **2018**, *359*, 926–930. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, P.; Song, Q.; Ren, J.; Zhang, W.; Wang, Y.; Zhou, L.; Wang, D.; Chen, K.; Jiang, L.; Zhang, B.; et al. Simultaneous analysis of mutations and methylations in circulating cell-free DNA for hepatocellular carcinoma detection. *Sci. Transl. Med.* **2022**, *14*, eabp8704. [\[CrossRef\]](#)
- Visal, T.H.; den Hollander, P.; Cristofanilli, M.; Mani, S.A. Circulating tumour cells in the -omics era: How far are we from achieving the 'singularity'? *Br. J. Cancer* **2022**, *127*, 173. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, K.; Sun, J.; Zhao, H.; Jiang, R.; Zheng, J.; Li, Z.; Peng, J.; Shen, H.; Zhang, K.; Zhao, J.; et al. Non-invasive lung cancer diagnosis and prognosis based on multi-analyte liquid biopsy. *Mol. Cancer* **2021**, *20*, 23. [\[CrossRef\]](#)
- Tomeva, E.; Switzeny, O.J.; Heitzinger, C.; Hippe, B.; Haslberger, A.G. Comprehensive Approach to Distinguish Patients with Solid Tumors from Healthy Controls by Combining Androgen Receptor Mutation p.H875Y with Cell-Free DNA Methylation and Circulating miRNAs. *Cancers* **2022**, *14*, 462. [\[CrossRef\]](#)
- Lin, A.A.; Nimgaonkar, V.; Issadore, D.; Carpenter, E.L. Extracellular Vesicle-Based Multianalyte Liquid Biopsy as a Diagnostic for Cancer. *Annu. Rev. Biomed. Data Sci.* **2022**, *5*, 269–292. [\[CrossRef\]](#)
- Li, L.; Ly, M.; Linhardt, R.J. Proteoglycan sequence. *Mol. Biosyst.* **2012**, *8*, 1613–1625. [\[CrossRef\]](#)
- Bülow, H.E.; Hobert, O. The Molecular Diversity of Glycosaminoglycans Shapes Animal Development. *Annu. Rev. Cell Dev. Biol.* **2006**, *22*, 375–407. [\[CrossRef\]](#)
- Gray, A.L.; Pun, N.; Ridley, A.J.L.; Dyer, D.P. Role of extracellular matrix proteoglycans in immune cell recruitment. *Int. J. Exp. Pathol.* **2022**, *103*, 34–43. [\[CrossRef\]](#) [\[PubMed\]](#)
- du Souich, P.; García, A.G.; Vergés, J.; Montell, E. Immunomodulatory and anti-inflammatory effects of chondroitin sulphate. *J. Cell. Mol. Med.* **2009**, *13*, 1451–1463. [\[CrossRef\]](#)
- Iozzo, R.V.; Schaefer, L. Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biol.* **2015**, *42*, 11–55. [\[CrossRef\]](#)
- Gesslbauer, B.; Rek, A.; Falsone, F.; Rajkovic, E.; Kungl, A.J. Proteoglycanomics: Tools to unravel the biological function of glycosaminoglycans. *Proteomics* **2007**, *7*, 2870–2880. [\[CrossRef\]](#) [\[PubMed\]](#)
- Weyers, A.; Yang, B.; Yoon, D.S.; Park, J.H.; Zhang, F.; Lee, K.B.; Linhardt, R.J. A Structural Analysis of Glycosaminoglycans from Lethal and Nonlethal Breast Cancer Tissues: Toward a Novel Class of Theragnostics for Personalized Medicine in Oncology? *OMICS* **2012**, *16*, 79–89. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gatto, F.; Bratulic, S.; Jonasch, E.; Limeta, A.; Maccari, F.; Galeotti, F.; Volpi, N.; Lundstam, S.; Nielsen, J.; Stierner, U. Plasma and Urine Free Glycosaminoglycans as Monitoring and Predictive Biomarkers in Metastatic Renal Cell Carcinoma: A Prospective Cohort Study. *JCO Precis. Oncol.* **2023**, *7*, e2200361. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gatto, F.; Dabestani, S.; Bratulic, S.; Limeta, A.; Maccari, F.; Galeotti, F.; Volpi, N.; Stierner, U.; Nielsen, J.; Lundstam, S. Plasma and Urine Free Glycosaminoglycans as Monitoring Biomarkers in Nonmetastatic Renal Cell Carcinoma—A Prospective Cohort Study. *Eur. Urol. Open Sci.* **2022**, *42*, 30–39. [\[CrossRef\]](#)
- Al-Nakouzi, N.; Wang, C.K.; Oo, H.Z.; Nelepçu, I.; Lallous, N.; Spliid, C.B.; Khazamipour, N.; Lo, J.; Truong, S.; Collins, C.; et al. Reformation of the chondroitin sulfate glycocalyx enables progression of AR-independent prostate cancer. *Nat. Commun.* **2022**, *13*, 4760. [\[CrossRef\]](#)
- Theocharis, A.D.; Vynios, D.H.; Papageorgakopoulou, N.; Skandalis, S.S.; Theocharis, D.A. Altered content composition and structure of glycosaminoglycans and proteoglycans in gastric carcinoma. *Int. J. Biochem. Cell Biol.* **2003**, *35*, 376–390. [\[CrossRef\]](#)
- Balbisi, M.; Sugár, S.; Schlosser, G.; Szeitz, B.; Fillinger, J.; Moldvay, J.; Drahoš, L.; Szász, A.M.; Tóth, G.; Turiák, L. Inter- and intratumoral proteomics and glycosaminoglycan characterization of ALK rearranged lung adenocarcinoma tissues: A pilot study. *Sci. Rep.* **2023**, *13*, 6268. [\[CrossRef\]](#)
- Li, G.; Li, L.; Joo, E.J.; Son, J.W.; Kim, Y.J.; Kang, J.K.; Lee, K.B.; Zhang, F.; Linhardt, R.J. Glycosaminoglycans and glycolipids as potential biomarkers in lung cancer. *Glycoconj. J.* **2017**, *34*, 661–669. [\[CrossRef\]](#) [\[PubMed\]](#)

22. Pál, D.; Tóth, G.; Sugár, S.; Fügedi, K.D.; Szabó, D.; Kovalszky, I.; Papp, D.; Schlosser, G.; Tóth, C.; Tornóczy, T.; et al. Compositional Analysis of Glycosaminoglycans in Different Lung Cancer Types—A Pilot Study. *Int. J. Mol. Sci.* **2023**, *24*, 7050. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Mattox, A.K.; Douville, C.; Wang, Y.; Popoli, M.; Ptak, J.; Silliman, N.; Dobbryn, L.; Schaefer, J.; Lu, S.; Pearlman, A.H.; et al. The Origin of Highly Elevated Cell-Free DNA in Healthy Individuals and Patients with Pancreatic, Colorectal, Lung, or Ovarian Cancer. *Cancer Discov.* **2023**, *13*, 2166–2179. [\[CrossRef\]](#)
24. Bratulic, S.; Limeta, A.; Dabestani, S.; Birgisson, H.; Enblad, G.; Ståhlberg, K.; Hesselager, G.; Häggman, M.; Höglund, M.; Simonson, O.E.; et al. Noninvasive detection of any-stage cancer using free glycosaminoglycans. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2115328119. [\[CrossRef\]](#)
25. Wang, F.; Li, X.; Li, M.; Liu, W.; Lu, L.; Li, Y.; Chen, X.; Yang, S.; Liu, T.; Cheng, W.; et al. Ultra-short cell-free DNA fragments enhance cancer early detection in a multi-analyte blood test combining mutation, protein and fragmentomics. *Clin. Chem. Lab. Med.* **2023**, *62*, 168–177. [\[CrossRef\]](#) [\[PubMed\]](#)
26. D’Ambrosi, S.; Giannoukakos, S.; Antunes-Ferreira, M.; Pedraz-Valdunciel, C.; Bracht, J.W.; Potie, N.; Gimenez-Capitan, A.; Hackenberg, M.; Fernandez Hilario, A.; Molina-Vila, M.A.; et al. Combinatorial Blood Platelets-Derived circRNA and mRNA Signature for Early-Stage Lung Cancer Detection. *Int. J. Mol. Sci.* **2023**, *24*, 4881. [\[CrossRef\]](#)
27. Bratulic, S.; Limeta, A.; Maccari, F.; Galeotti, F.; Volpi, N.; Levin, M.; Nielsen, J.; Gatto, F. Analysis of normal levels of free glycosaminoglycans in urine and plasma in adults. *J. Biol. Chem.* **2022**, *298*, 101575. [\[CrossRef\]](#)
28. da Silva, M.N.R.; Mendes, A.; Martins, J.R.M.; Tobias-Machado, M.; Pinhal, M.A.d.S. Prospective Evaluation of Chondroitin Sulfate, Heparan Sulfate and Hyaluronic Acid in Prostate Cancer. *Int. Braz. J. Urol.* **2018**, *44*, 1139–1146. [\[CrossRef\]](#)
29. Saito, A.; Munakata, H. Analysis of plasma proteins that bind to glycosaminoglycans. *Biochim. Biophys. Acta (BBA)-Gen. Subj.* **2007**, *1770*, 241–246. [\[CrossRef\]](#)
30. Prabhakar, V.; Sasisekharan, R. The biosynthesis and catabolism of galactosaminoglycans. *Adv. Pharmacol.* **2006**, *53*, 69–115.
31. Cohen, J.F.; Korevaar, D.; Altman, D.G.; Bruns, D.; Gatsonis, C.; Hooft, L.; Irwig, L.; Levine, D.; Reitsma, J.B.; de Vet, H.C.W.; et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open* **2016**, *6*, e012799. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Detterbeck, F.C.; Boffa, D.J.; Kim, A.W.; Tanoue, L.T. The Eighth Edition Lung Cancer Stage Classification. *Chest* **2017**, *151*, 193–203. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Travis, W.D.; Brambilla, E.; Müller-Hermelink, H.K.; Harris, C.C. *Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart*; IARC Press: Lyon, France, 2004.
34. Tamburro, D.; Bratulic, S.; Abou Shameh, S.; Soni, N.K.; Bacconi, A.; Maccari, F.; Galeotti, F.; Mattsson, K.; Volpi, N.; Nielsen, J.; et al. Analytical performance of a standardized kit for mass spectrometry-based measurements of human glycosaminoglycans. *J. Chromatogr. B* **2021**, *1177*, 122761. [\[CrossRef\]](#)
35. Volpi, N.; Galeotti, F.; Yang, B.; Linhardt, R.J. Analysis of glycosaminoglycan-derived, precolumn, 2-aminoacridone-labeled disaccharides with LC-fluorescence and LC-MS detection. *Nat. Protoc.* **2014**, *9*, 541–558. [\[CrossRef\]](#)
36. Qvick, A.; Stenmark, B.; Carlsson, J.; Isaksson, J.; Karlsson, C.; Helenius, G. Liquid biopsy as an option for predictive testing and prognosis in patients with lung cancer. *Mol. Med.* **2021**, *27*, 68. [\[CrossRef\]](#)
37. Kruschke, J.K. Bayesian estimation supersedes the *t* test. *J. Exp. Psychol. Gen.* **2013**, *142*, 573–603. [\[CrossRef\]](#)
38. Vehtari, A.; Gelman, A.; Simpson, D.; Carpenter, B.; Bürkner, P.-C. Rank-Normalization, Folding, and Localization: An Improved  $\hat{R}$  for Assessing Convergence of MCMC (with Discussion). *Bayesian Anal.* **2021**, *16*, 667–718. [\[CrossRef\]](#)
39. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* **2017**, *80*, 1–28. [\[CrossRef\]](#)
40. Makowski, D.; Ben-Shachar, M.S.; Lüdtke, D. bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *J. Open Source Softw.* **2019**, *4*, 1541. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.