

1 Plasma glycosaminoglycans and cell-free DNA to discriminate benign 2 and malignant lung diseases

3 Running title: Glycosaminoglycans & cfDNA in Lung Cancer

4 Family name followed by given name: Qvick Alvida¹, Bratulic Sinisa², Carlsson Jessica³,

5 Stenmark Bianca¹, Karlsson Christina⁴, Nielsen Jens^{2,5}, Gatto Francesco^{2,6}, Helenius Gisela¹

6 Affiliations

7 1) Dept. of Laboratory Medicine, Faculty of Medicine and Health, Örebro
8 University, 701 82 Örebro, Sweden

9 2) Department of Life Sciences, Chalmers University of Technology, 412 96
10 Gothenburg, Sweden

11 3) Dept. of Urology, Faculty of Medicine and Health, Örebro University, 701 82
12 Örebro, Sweden

13 4) School of Health Sciences, Örebro University, 701 82 Örebro, Sweden

14 5) BioInnovation Institute, Ole Maaløes Vej 3, 2200 Copenhagen, Denmark

15 6) Department of Oncology-Pathology, Karolinska Institute, 171 77 Stockholm,
16 Sweden

17

18 Correspondence to: Alvida Qvick

19 Email: alvida.qvick@regionorebrolan.se

20 Adress: Dep. of Laboratory Medicine, Örebro University Hospital,

21 Södra Grev. Roseng., 701 85 Örebro, Sweden

22 Bianca Stenmark: bianca.stenmark@regionorebrolan.se

23 Jessica Carlsson: jessica.carlsson@regionorebrolan.se

24 Sinisa Bratulic: bratulic@chalmers.se, sinisa.bratulic@elypta.com

25 Christina Karlsson: Christina.Karlsson@oru.se

26 Francesco Gatto: francesco.gatto@ki.se

27 Jens Nielsen: nielsenj@chalmers.se

28 Gisela Helenius: gisela.helenius@regionorebrolan.se

29 All authors have read and approved the final version of this manuscript.

30 **Conflict of interest:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

31 F. Gatto and J. Nielsen are shareholders in Elypta AB. F. Gatto and S. Bratulic are employed at
32 Elypta AB. J. Nielsen is a board member at Elypta AB. Elypta AB has a commercial interest in
33 part of the technology described in this study. Remaining authors have no conflicts to declare.

34 1. Abstract

35 We aimed to investigate the use of free glycosaminoglycan profiles (GAGomes) and cfDNA
36 in plasma to differentiate between lung cancer and benign lung disease. GAGs were analyzed
37 using the MIRAM® Free Glycosaminoglycan Kit with ultra-high-performance liquid
38 chromatography and electrospray ionization triple-quadrupole mass spectrometry. We
39 detected two GAGome features, 0S chondroitin sulfate (CS) and 4S CS, with cancer-specific
40 changes. Based on the observed GAGome changes, we devised a model to predict lung
41 cancer. The model, named the GAGome score, could detect lung cancer with 41.2%
42 sensitivity (95% CI: 9.2-54.2%) at 96.4% specificity (CI: 95.2-100.0%, n=113). Furthermore,
43 we found that the GAGome score, when combined with a cfDNA test, could increase the
44 sensitivity for lung cancer from 42.6% (95% CI: 31.7-60.6%, cfDNA alone) to 70.5% (CI:
45 57.4 - 81.5%) at 95% specificity (CI: 75.1-100%, n=74). Notably, the combined GAGome
46 and cfDNA testing improved the sensitivity, especially in early stages, relative to the cfDNA
47 alone. Our findings show that plasma GAGome profiles can enhance cfDNA testing
48 performance, highlighting the applicability of a multiomics approach in lung cancer
49 diagnostics.

50

51 Keywords: cfDNA/GAGome/glycosaminoglycans/lung cancer/multiomics

52

53 2. Introduction

54 The use of high-throughput sequencing to analyze the genome and its aberrations has
55 advanced diagnostics, predictive testing, and monitoring for many types of cancer. To avoid
56 the challenges of obtaining tissue biopsies for cancers like lung cancer (LC), a liquid biopsy
57 has emerged as a promising alternative. Numerous articles have been published on sequencing
58 of circulating cell-free DNA (cfDNA) for such cancers¹. Despite technical advances, which
59 improved the sensitivity of genomics-only liquid biopsy assays, many cancers remain
60 undetected. This can be due to technical as well as biological reasons since the released
61 amount of circulating tumor DNA is very heterogeneous between different tumors². To
62 overcome this limitation, the field has turned beyond genomics to a combination of different
63 omics techniques, referred to as multiomics, including combined information from genomics,
64 transcriptomics, proteomics, methylomics, metabolomics, extracellular vesicles, and
65 circulating tumor cells among others³⁻⁸.

66 Glycosaminoglycans (GAGs) are unbranched linear polysaccharides that can be divided into
67 four main classes: chondroitin sulfate, heparin sulfate, keratan sulfate, and hyaluronic acid⁹.
68 The sulfation and glycosylation patterns in GAG chains can vary widely, leading to a very
69 high structural and functional diversity of resulting GAG molecules¹⁰. GAGs play crucial
70 roles in cellular functions, including maintaining the extracellular matrix structure and
71 providing hydration to cells. They are also integral to the immune response^{11,12}, tissue
72 homeostasis, as well as cell growth, proliferation, differentiation, and adhesion¹³. Notably,
73 GAGs have been implicated in different aspects of cancer development and progression due
74 to their involvement in the tumor microenvironment, through their interaction with growth
75 factors, growth factor receptors, and cytokines¹⁴. The disaccharide composition of GAGs in
76 tumor tissue, plasma, and urine has been shown to be altered in several cancer types such as
77 breast, prostate, gastric, and renal cell carcinoma¹⁵⁻¹⁹. Conversely, only a few studies

78 measured the structural profile of GAG disaccharides (or GAGome) in tissue or liquid
79 biopsies in LC^{20,21}, indicating the need for further research.

80 This study aimed to explore the plasma GAGome's potential for differentiating LC from non-
81 malignant lung diseases via liquid biopsy. Additionally, we investigated if combining
82 GAGome and cfDNA data could increase the sensitivity and specificity to detect LC.

83 3. Materials and Methods

84 3.1 Study design and cohort characteristics

85 This study was reported in compliance with the Standards for Reporting of Diagnostic
86 Accuracy (STARD) guidelines²² (Table S1). Study participants were enrolled into the study
87 during routine clinical investigation, at the lung clinic at Örebro University Hospital, between
88 February 2016 and February 2019. Patients investigated for the suspicion of LC were
89 included in the study and formed a consecutive series. To ensure the most accurate diagnosis,
90 all LC cases were histologically confirmed on tumor tissue. Patients diagnosed with cancer of
91 other origin than lung, sample not collected prior to treatment start, or with inadequate tumor
92 material for diagnosis, were excluded from the study. Participants gave written informed
93 consent before inclusion and the study was approved by the regional ethics committee board
94 in Uppsala (Approval 2015-400, 2021-01478).

95 Tumors were staged and histologically classified according to the guidelines of the
96 International Association for the Study of Lung Cancer (IASLC) and the World Health
97 Organization nomenclature, respectively^{23,24}. Patients with benign lung diseases mainly
98 consisted of different pulmonary obstructive diseases, inflammatory conditions in the lung,
99 fibrosis in the lung or benign lung nodules.

100 3.2 Plasma collection and isolation

101 Blood was collected in Cell-Free RNA™ BCT tubes (Streck, Omaha, NE) and plasma was
102 retrieved by a two-step centrifugation; 2000×g for 10 min followed by 16 000×g for 10 min.
103 Plasma was stored at -80° C until preparation for analysis. Samples were thawed on ice and an
104 aliquot of each sample was taken for further analyses.

105 3.3 GAGome measurements

106 Plasma GAGome measurements were performed retrospectively in a single blinded Good
107 Laboratory Practice (GLP)-compliant central laboratory using MIRAM® Free
108 Glycosaminoglycan Kit (Elypta AB, Sweden), which is a standardized kit for GAG
109 extraction, detection, and quantification by ultra-high-performance liquid chromatography
110 (UHPLC) coupled with electrospray ionization triple-quadrupole mass spectrometry system
111 (ESI-MS/MS, Acquity I-class Plus Xevo TQ-S micro, Waters® Corporation, MA). A single
112 UHPLC column equipped with a pre-column guard (Waters® ACQUITY UPLC BEH C18
113 VanGuard Pre-column) was sufficient to analyze all samples in this study with no quality
114 deterioration observed over time. The analytical performance characteristics of the kit have
115 been previously described²⁵.

116 In short, the kit is based on a method by Volpi *et al.* (2014)²⁶. The assay consists of the
117 enzymatic depolymerization of GAGs from the sample into disaccharides by *Chondroitinase*
118 *ABC* and *Heparinase I-II-III*. The method omits proteolytic digestion, thereby limiting the
119 derived depolymerized GAGs to the protein-free fraction – or free GAGs. Following
120 depolymerization, disaccharides are labeled using 2-aminoacridone and injected into an
121 UHPLC-MS/MS for separation and detection. The peaks of the 17 disaccharides are acquired
122 at using multiple reaction monitoring analysis implemented in the mass spectrometry software
123 (Waters® TargetLynx). The chromatographic conditions and MS configuration were set in
124 accordance with the kit instruction for use.

125 Each sample was measured in singleton. The so-measured GAGome consisted of the absolute
126 concentrations of 17 GAG disaccharides, corresponding to eight different sulfation patterns of
127 chondroitin sulfate (CS) and heparan sulfate (HS), and one hyaluronic acid (HA)
128 disaccharide. Specifically, we quantified eight CS disaccharides (0S CS, 2S CS, 6S CS, 4S
129 CS, 2S6S CS, 2S4S CS, 4S6S CS, and TriS CS) and eight HS disaccharides (0S HS, 2S HS,
130 6S HS, NS HS, NS6S HS, NS2S HS, 2S6S HS, and TriS HS). We expanded the GAGome to
131 include an additional 22 calculated features informative of GAG biology: a) the total CS and
132 total HS concentration as the sum of the corresponding disaccharide concentrations, b) the CS
133 charge [-] and HS charge [-] as the weighted sum of sulfated disaccharides, where the weight
134 is the count of sulfo- groups in each disaccharide, c) two ratios (4S CS/0S CS and 6S CS/0S
135 CS), and d) the relative concentration (or mass fraction, in $\mu\text{g}/\mu\text{g} \%$) of each of the 16 CS and
136 HS disaccharides by normalizing each absolute concentration by the total CS and HS
137 concentration, respectively. For each sample, the GAGome consisted of maximally 39
138 features.

139 We considered a GAGome feature detectable in plasma if the median concentration across all
140 samples was above $0.1 \mu\text{g}/\text{mL}^{25}$. GAGome features that did not fulfill this criterion were
141 excluded from downstream analyses.

142 3.4 cfDNA data

143 For 81 of the 113 samples, cfDNA concentration measurement was available. The cfDNA
144 was extracted from 4 mL plasma using the QIASymphony DSP Circulating DNA kit on the
145 QIASymphony SP system (Qiagen, Germany) according to the manufacturer's instructions.
146 The concentration was measured using dsDNA HS assay kit (Thermo Fisher) on Qubit 2.0
147 Fluorometer (Thermo Fisher). Out of the samples with available cfDNA concentration
148 measurements, 74 samples also had NGS data available, which have been published

149 previously²⁷. The NGS panel used was AVENIO ctDNA Surveillance kit (Roche Diagnostics,
150 Rotkreuz, Switzerland), which includes 197 cancer relevant genes.

151 3.5 Statistical analysis

152 Continuous data were presented using mean and median values while categorical data were
153 presented using absolute and relative frequencies. Differences between LC patients and controls
154 regarding general clinical characteristics were investigated using χ^2 -tests for categorical
155 variables and Student's t-tests for continuous data.

156 3.5.1 GAGome analysis

157 We compared levels of each detectable GAGome feature in cancer versus control participants
158 using Bayesian estimation with practical equivalence testing²⁸. First, we fitted a Bayesian linear
159 regression where each individual GAGome feature was standardized and modelled as a
160 normally distributed response variable and the disease state (case vs control) was the only
161 binary predictor. Second, we controlled the correlation between GAGome measurements in
162 cancer vs controls for technical variation by fitting Bayesian linear models as above but
163 including experimental batch (binary) and sample age (continuous, in months) as predictors.
164 We used a t-distribution centered on 0 with 7 degrees of freedom, and scale = 2 for all the priors
165 in all the Bayesian models. We fit the models using *rstanarm* package (ver 2.21.3) in R (ver.
166 4.2.1). The convergence and stability of the Bayesian sampling was assessed using R-hat, which
167 should be below 1.01²⁹ and Effective Sample Size (ESS), which should be greater than 1000³⁰.
168 The same convergence criteria were used for all Bayesian models.

169 A GAGome feature was considered credibly associated to case-control status if the 95%
170 confidence interval (CI) of the difference in means did not include zero, and no more than 10%
171 of the CI passed into the Region of Practical Equivalence (ROPE)²⁸. The ROPE-interval was
172 defined as [-0.1, 0.1] of the standardized mean, corresponding to a negligible effect size³¹.
173 GAGome features that were credibly associated with case-control status after adjusting the

174 linear regression model for experimental batch and sample age were further analyzed as
175 predictors of cancer in the GAGome score development.

176 3.5.2 GAGome score development

177 A Bayesian logistic model was fitted to predict case vs control using the absolute concentration
178 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
179 batch-normalized before fitting the model. The Markov chain Monte Carlo (MCMC) sampling
180 was performed with four chains of 10 000 iterations and a warmup of 5000. The output of the
181 model is referred to as *plasma GAGome LC score*, or simply *GAGome score*.

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

186 3.5.3 cfDNA score development

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

190 3.5.4 Combined GAGome and cfDNA test pathway

191 A combined test was envisioned as a diagnostic pathway that sequentially uses the cfDNA score
192 and the GAGome score to render a diagnostic decision, i.e. “combined test positive” vs
193 “combined test negative”. To cumulatively retain the 95% specificity for the combined test, as
194 was rationalized for the GAGome score, the GAGome specificity was kept at 95% while the
195 specificity for the cfDNA score was set to 100%. Specifically, the GAGome score, when
196 positive, could be used to reclassify cfDNA score false negative samples as “combined test
197 positive”. The testing procedure was as following: 1) calculate the cfDNA score cut-off at
198 which a minimum of 100% specificity was achieved, 2) calculate the GAGome score cut-off at

199 which a minimum of 95% specificity was achieved, 3) for each sample, consider it positive for
200 the cfDNA score, if the sample score was above the cut-off; analogously for the GAGome score,
201 4) for each sample, consider it positive for the combined test if at least one between the cfDNA
202 and GAGome score were positive, 5) mark samples which were cfDNA score-negative but
203 GAGome score-positive as “reclassified positive”.

204 4. Results

205 4.1 Cohort characteristics

206 The cohort comprised patients referred to the clinic under suspicion of LC (n=113). Of these,
207 85 cases were subsequently diagnosed with LC, which were further stratified into non-small
208 cell LC (NSCLC, n=77) and small cell LC (SCLC, n=8). Patients diagnosed with a benign
209 lung disease (n=28) served as controls (Table 1). Cases were slightly older than controls
210 (mean 70.4 vs 67.0 years) and more evenly distributed among the sexes (52.9% females vs
211 60.7% females), neither characteristic showed a significant difference. A history of smoking
212 was more prevalent among cases compared to controls (85.9% vs 60.8%, $p=0.047$). The
213 predominant tumor stage was stage IV (61.2%), with a mean tumor size of 4.48 cm.

214 **Table 1.** Clinical characteristics of the cohort. Cases include patients diagnosed with lung
215 cancer, irrespective of histology, and controls include patients diagnosed with different benign
216 diseases of the lung.

	Case (n=85)	Control (n=28)	Lung cancer vs Controls, p -value	Overall (N=113)
Age				
Mean (SD)	70.4 (8.5)	67.0 (14.4)	0.246	69.6 (10.3)
Gender				
Female	45 (52.9%)	17 (60.7%)	0.473	63 (55.3%)
Male	40 (47.1%)	11 (39.3%)		51 (44.7%)
Smoking				
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%)

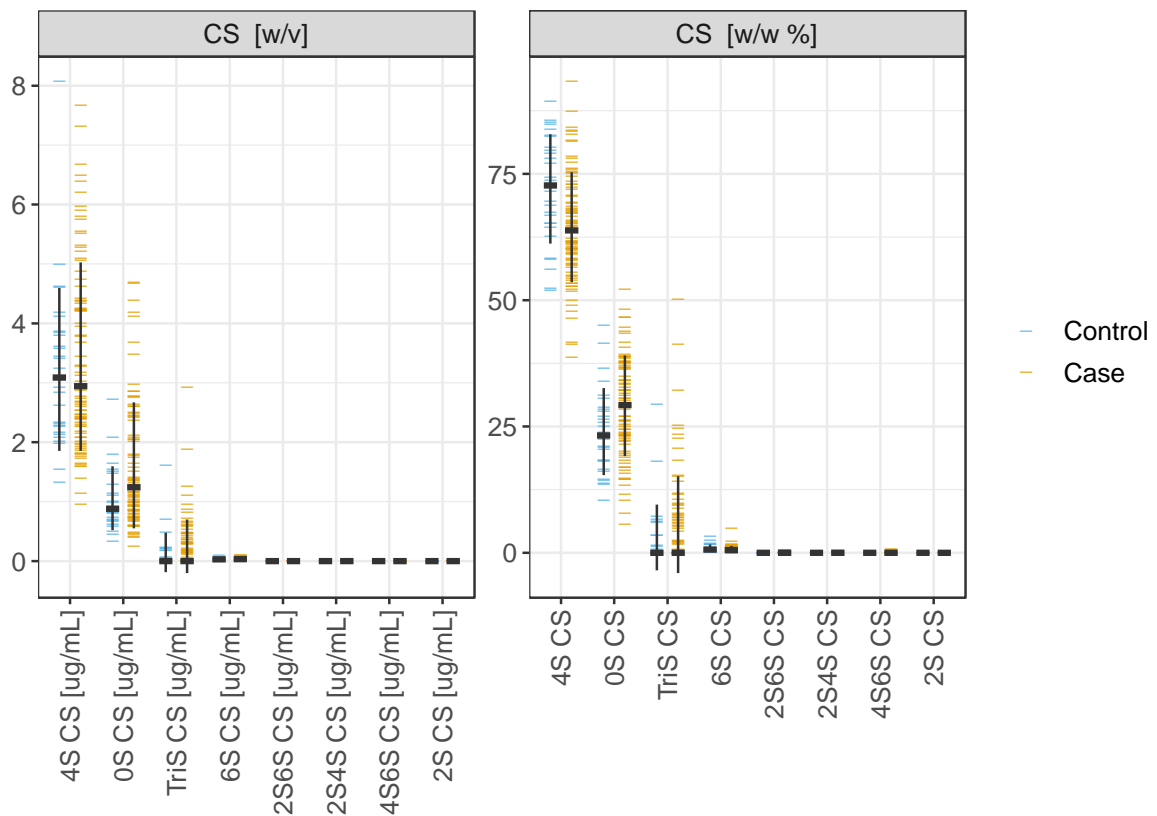
	Case (n=85)	Control (n=28)	Lung cancer vs Controls, <i>p</i> -value	Overall (N=113)
Tumor size (cm)				
Mean (SD)	4.5 (2.4)	NA (NA)		4.5 (2.4)
Median [Min, Max]	4.0 [0.9, 12.5]	NA [NA, NA]		4.0 [0.9, 12.5]
Missing	4 (4.7%)	28 (100%)		32 (28.3%)
IASLC stage				
I	9 (10.6%)	0 (0%)		9 (8.0%)
II	6 (7.1%)	0 (0%)		6 (5.3%)
III	18 (21.2%)	0 (0%)		18 (15.9%)
IV	52 (61.2%)	0 (0%)		52 (46.0%)

217 SD = Standard deviation, NA = Not Applicable and IASLC = International association for the study of lung
218 cancer.

219 4.2 Correlation between plasma GAGomes and lung cancer diagnosis

220 The plasma GAGome, which comprised the structural characterization of 17 different GAG
221 disaccharides, of all samples (n=113) was analyzed successfully. 4S CS and 0S CS were the
222 only GAG disaccharides with a median detected concentration above 0.1 µg/mL. Therefore,
223 we focused the subsequent analysis on these two independently measured features and the
224 corresponding four derived features (Materials and methods, table S2, figure 1).

225 Compared to controls, cases with cancer had a nominally higher median concentration of 0S
226 CS, and a lower concentration of 4S CS (figure 1, table S2). Concomitantly, we found that the
227 composition of CS in cases with cancer shifted towards a lower 4S CS fraction and higher 0S
228 CS fraction (figure 1).



229

230 **Figure 1.** Plasma GAGome profiles. Concentrations (left) and fractions (right) of plasma
231 GAGs in cases (n=85) and controls (n=28). CS: chondroitin sulfate.

232

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

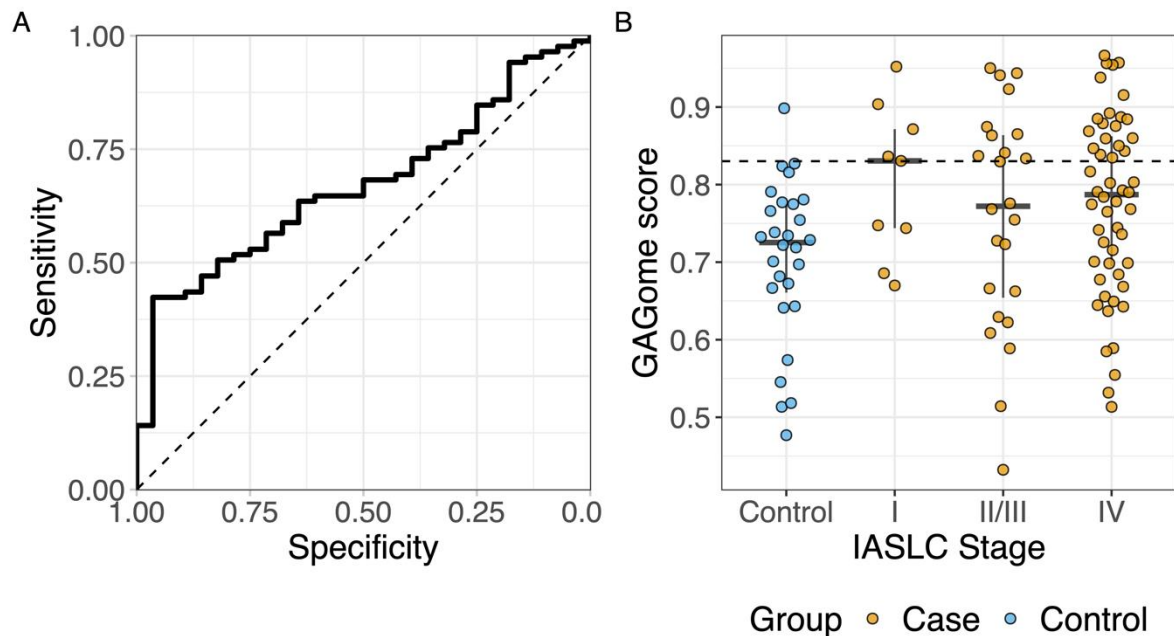
238

4.3 GAGome and cfDNA score

239

240 We next sought to develop a *plasma GAGome LC score* (or simply, GAGome score) for
241 discriminating LC cases and controls. We fitted a Bayesian logistic model to predict cancer
242 (figure S2, supplementary text), where the model's output, referred to as the GAGome score,
242 corresponded to the log-odds ratio of having LC.

243 The GAGome score had a 41.2% sensitivity (95% CI: 9.2-54.2%) to detect LC at a 96.4%
244 specificity (95% CI: 95.2-100%) with an AUC of 0.67 (95% CI: 0.56-0.77, figure 2, table S3).



245 **Figure 2.** Performance of the GAGome score to discriminate cases and controls in A) the
246 whole cohort and B) by IASLC stage. Controls are shown in blue ($N_{\text{Control}} = 28$) and cases in
247 orange ($N_{\text{Case}} = 85$; $N_{\text{Stage I}} = 9$, $N_{\text{Stage II/III}} = 24$, $N_{\text{Stage IV}} = 58$). IASLC: International Association
248 for the Study of Lung Cancer.
249

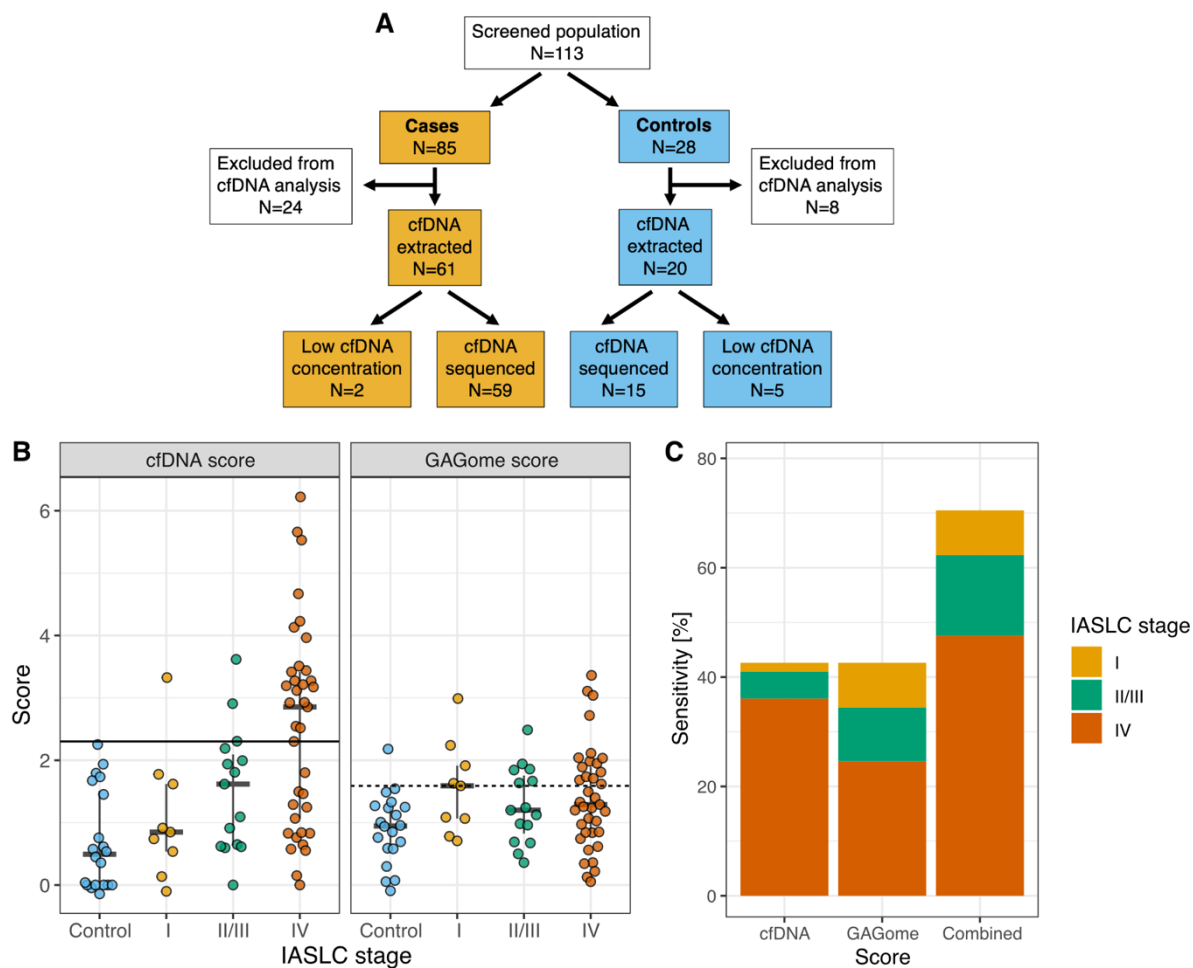
250

251 Next, cfDNA data was analyzed. Neither cfDNA concentration nor number of cfDNA
252 variants were correlated to the GAGome features included in the GAGome score (OS CS
253 concentration and 4S CS fraction, figure S3). A *cfDNA score* was generated in the subset
254 samples for which both cfDNA concentration and number of variants were available ($n=74$,
255 figure 3A). Specifically, we fitted a Bayesian logistic model to predict case vs control using
256 cfDNA concentration and the number of cfDNA variants as predictors. The model's output,
257 referred to as the *cfDNA score*, corresponded to the log-odds ratio of having LC. We treated
258 samples where cfDNA concentration was available but insufficient for cfDNA variant

259 analysis (n=7) as cfDNA score negative for evaluating the cfDNA score performance. The
260 cfDNA score had an AUC of 0.80 (95% CI: 0.685-0.903) and a 42.6% sensitivity (95% CI:
261 31.7-60.6%) at 100% specificity (figure 3B).

262 We envisioned a diagnostic pathway that relies on sequential cfDNA and GAGome
263 measurements in plasma, specifically to use the GAGome score for potentially cfDNA score-
264 false negatives. This combined test would be positive if either the cfDNA or GAGome scores
265 were positive. We allowed for one false positive and set the threshold for cfDNA positivity at
266 100% specificity and GAGome score positivity at 95% specificity (figure 3). The sensitivity
267 for the combined test increased to 70.5% (95% CI: 57.4 - 81.5%) at a specificity of 95% (95%
268 CI: 75.1-100%, figure 3C). Notably, in this diagnostic pathway with overall 95% specificity
269 (one false positive), the GAGome score contributed to a higher sensitivity for stage I (55.6%
270 vs 11.1% for cfDNA alone, table S4) and stage II-III (40% vs 20% for cfDNA alone, table
271 S4), and could correctly reclassify 17 out of 19 cases that were false negative when using the
272 cfDNA score alone (table S5).

273



274
 275 **Figure 3.** Performance of the combined test. A) Subject flow and inclusion for the cfDNA
 276 data. B) Performance and cutoffs for the cfDNA and GAGome score separately and C)
 277 sensitivity when combined (by IASCL stage). IASCL: International Association for the Study
 278 of Lung Cancer.

279 5. Discussion

280 In this study, we aimed to use GAGomes to discriminate between LC and benign lung
 281 diseases and to explore if the addition of cfDNA data could further increase the sensitivity of
 282 the developed score. We used the detected features, 0S CS and 4S CS, to build a GAGome
 283 score that reached a 41.2% sensitivity at 96.4% specificity. We next envisioned a test that
 284 combines the GAGome-score with cfDNA data to increase the diagnostic potential for LC. In
 285 the subset of patients with cfDNA data available the combined test could diagnose LC with
 286 70.5% sensitivity at 95% specificity. To the best of our knowledge, this is the first study to
 287 develop a multiomics test combining GAGs and cfDNA for cancer.

288 Research on GAGs in LC has been limited, with most studies focusing on tissue-based
289 analysis. An increase in total CS has been observed in tumor tissue compared to normal lung
290 tissue^{20,21,32}. However, the results regarding CS sulfation patterns have been inconsistent. Pál
291 *et al.* and Balbisi *et al.* reported lower 0S CS and higher 4S CS in tumors, while Li *et al.*
292 found the opposite for 4S CS. Discrepancies between tissue and plasma can be explained by
293 the cell origin of the molecules analyzed. Mattox *et al.* investigated the contribution of
294 different cells to cfDNA in plasma and found that, even in cancer patients, over 70% of
295 cfDNA originated from leukocytes, and only 2.2% of the fragments in LC patients originated
296 from the lung³³. They concluded that these results reflect the systemic effect of the tumor on
297 the body, suggesting that what is detected is not primarily the tumor itself, but the effect it
298 causes. This concept can be applied to other circulating biomarkers, such as GAGs, indicating
299 that plasma tests should be interpreted separately from tissue tests. Circulating GAGs in
300 plasma has only been reported once previously³⁴, where elevated 0S CS concentration and a
301 lower 4S CS fraction was detected in LC cases relative to controls.

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

309 The effectiveness of using multiomics for cancer detection was notably proposed by Cohen *et al.*
310 *al.* with CancerSEEK, which analyzed mutations in cfDNA and protein levels³. They reported
311 an overall sensitivity of 70% across all tumor types and about 59% for LC specifically. Since
312 then, several other studies have explored the potential of multiomics as a diagnostic tool.

313 For instance, Wang *et al.*⁴ tested a combination of cfDNA variants, proteins and
314 fragmentomics on a cohort of colorectal, esophageal, gastric, liver, lung, and ovarian cancers
315 ³⁵. At 98% specificity, cfDNA variants alone showed a sensitivity of 46%. This increased to
316 60% with the addition of proteins, and further to 66% with fragmentomics. However, for LC
317 alone, the sensitivity of the combined model only reached 38.5%. Chen *et al.* combined a
318 mutational score, methylation, and serum CEA levels for distinguishing between LC and
319 benign lung nodules, achieving a 76.9% sensitivity at a modest specificity of 58.3%. At the
320 95% specificity level used in our study, their sensitivity dropped to around 30%⁶. D'Ambrosi
321 *et al.* reported a promising 68% sensitivity at 95% specificity for diagnosing LC using
322 platelet-derived circRNA and mRNA³⁶. Notably, their control group primarily consisted of
323 asymptomatic individuals, similar to Wang *et al.* and CancerSEEK. Our study differed in
324 design by using patients referred to the clinic with lung-related symptoms qualifying for LC
325 evaluation, rather than selecting a specific control group. This enhances the clinical relevance
326 of our findings, as it better reflects real-world diagnostic challenges. Although our control
327 cohort was more suitable compared to several other multiomics studies, it is important to note
328 that most of these studies had a skew towards lower stages of LC, which may have negatively
329 impacted their reported sensitivity.

330 The only detectable plasma GAGome features in this cohort were 0S CS and 4S CS. This is
331 consistent with previous findings using this kit, both in healthy individuals and cancer^{34,37}.
332 Other studies have detected additional GAGs, such as HA in plasma³⁸. However, the methods
333 used in those studies included proteolytic digestion, while the method used here is
334 degradation-free, detecting only GAGs that circulate freely in the analyzed liquid. Saito *et al.*
335 tested the affinity of plasmatic proteins to different GAGs and could show that approximately
336 7.5% of the proteins are bound to HS and dermatan sulfate, while only 0.25% are bound to
337 CS³⁹. This suggests that CS has a higher probability of circulating freely in plasma,
338 supporting our findings.

339 A limitation of our study includes the relatively small sample size and lack of external cohorts
340 for validation. A methodological limitation is that blood was collected in tubes optimized for
341 stabilization of cells and extracellular RNA, and not immediately processed, as the
342 stabilization reagents are effective for several days. However, GAGs are primarily degraded
343 by highly substrate-specific enzymes in the lysosome⁴⁰. Since the tubes' main function is to
344 stabilize cells, lysosomal activity in the plasma fraction of the sample is unlikely. While an
345 unspecific degradation in plasma cannot be ruled out, the cancer-specific alterations to
346 GAGomes were robust to adjustment for sample age.

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

356 Funding

357 This work was funded by Nyckelfonden-Örebro University Hospital Research Foundation, Lions fund
358 for cancer research Uppsala-Örebro and Uppsala-Örebro Regional research council.

359 Data availability

360 Data is available from the authors upon reasonable request.

361

362 6. References

- 363 1. Chen, Z. *et al.* Circulating cell-free DNA as a diagnostic and prognostic biomarker for
364 non-small-cell lung cancer: a systematic review and meta-analysis. *Biomarkers in*
365 *Medicine* **14**, 587–597 (2020).
- 366 2. Bettegowda, C. *et al.* Detection of Circulating Tumor DNA in Early- and Late-Stage
367 Human Malignancies. *Science Translational Medicine* **6**, 224ra24-224ra24 (2014).
- 368 3. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-
369 analyte blood test. *Science* **359**, 926–930 (2018).
- 370 4. Wang, P. *et al.* Simultaneous analysis of mutations and methylations in circulating cell-
371 free DNA for hepatocellular carcinoma detection. *Science Translational Medicine* **14**,
372 eabp8704 (2022).
- 373 5. Visal, T. H., Hollander, P. den, Cristofanilli, M. & Mani, S. A. Circulating tumour cells in
374 the -omics era: how far are we from achieving the ‘singularity’? *British Journal of Cancer*
375 **127**, 173 (2022).
- 376 6. Chen, K. *et al.* Non-invasive lung cancer diagnosis and prognosis based on multi-analyte
377 liquid biopsy. *Molecular Cancer* **20**, 23 (2021).
- 378 7. Tomeva, E., Switzeny, O. J., Heitzinger, C., Hippe, B. & Haslberger, A. G.
379 Comprehensive Approach to Distinguish Patients with Solid Tumors from Healthy
380 Controls by Combining Androgen Receptor Mutation p.H875Y with Cell-Free DNA
381 Methylation and Circulating miRNAs. *Cancers (Basel)* **14**, 462 (2022).
- 382 8. Lin, A. A., Nimgaonkar, V., Issadore, D. & Carpenter, E. L. Extracellular Vesicle-Based
383 Multianalyte Liquid Biopsy as a Diagnostic for Cancer. *Annu. Rev. Biomed. Data Sci.* **5**,
384 269–292 (2022).
- 385 9. Li, L., Ly, M. & Linhardt, R. J. Proteoglycan sequence. *Mol. BioSyst.* **8**, 1613–1625
386 (2012).

- 387 10. Bülow, H. E. & Hobert, O. The Molecular Diversity of Glycosaminoglycans Shapes
388 Animal Development. *Annual Review of Cell and Developmental Biology* **22**, 375–407
389 (2006).
- 390 11. Gray, A. L., Pun, N., Ridley, A. J. L. & Dyer, D. P. Role of extracellular matrix
391 proteoglycans in immune cell recruitment. *International Journal of Experimental*
392 *Pathology* **103**, 34–43 (2022).
- 393 12. du Souich, P., García, A. G., Vergés, J. & Montell, E. Immunomodulatory and anti-
394 inflammatory effects of chondroitin sulphate. *J Cell Mol Med* **13**, 1451–1463 (2009).
- 395 13. Iozzo, R. V. & Schaefer, L. Proteoglycan form and function: A comprehensive
396 nomenclature of proteoglycans. *Matrix Biol* **42**, 11–55 (2015).
- 397 14. Gesslbauer, B., Rek, A., Falsone, F., Rajkovic, E. & Kungl, A. J. Proteoglycanomics:
398 tools to unravel the biological function of glycosaminoglycans. *Proteomics* **7**, 2870–2880
399 (2007).
- 400 15. Weyers, A. *et al.* A Structural Analysis of Glycosaminoglycans from Lethal and
401 Nonlethal Breast Cancer Tissues: Toward a Novel Class of Theragnostics for
402 Personalized Medicine in Oncology? *OMICS* **16**, 79–89 (2012).
- 403 16. Gatto, F. *et al.* Plasma and Urine Free Glycosaminoglycans as Monitoring and Predictive
404 Biomarkers in Metastatic Renal Cell Carcinoma: A Prospective Cohort Study. *JCO*
405 *Precision Oncology* **7**, e2200361 (2023).
- 406 17. Gatto, F. *et al.* Plasma and Urine Free Glycosaminoglycans as Monitoring Biomarkers in
407 Nonmetastatic Renal Cell Carcinoma-A Prospective Cohort Study. *Eur Urol Open Sci* **42**,
408 30–39 (2022).
- 409 18. Al-Nakouzi, N. *et al.* Reformation of the chondroitin sulfate glycocalyx enables
410 progression of AR-independent prostate cancer. *Nat Commun* **13**, 4760 (2022).
- 411 19. Theocharis, A. D., Vynios, D. H., Papageorgakopoulou, N., Skandalis, S. S. &
412 Theocharis, D. A. Altered content composition and structure of glycosaminoglycans and

- 413 proteoglycans in gastric carcinoma. *The International Journal of Biochemistry & Cell*
414 *Biology* **35**, 376–390 (2003).
- 415 20. Balbisi, M. *et al.* Inter- and intratumoral proteomics and glycosaminoglycan
416 characterization of ALK rearranged lung adenocarcinoma tissues: a pilot study. *Sci Rep*
417 **13**, 6268 (2023).
- 418 21. Li, G. *et al.* Glycosaminoglycans and glycolipids as potential biomarkers in lung cancer.
419 *Glycoconj J* **34**, 661–669 (2017).
- 420 22. Cohen, J. F. *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies:
421 explanation and elaboration. *BMJ Open* **6**, e012799 (2016).
- 422 23. Detterbeck, F. C., Boffa, D. J., Kim, A. W. & Tanoue, L. T. The Eighth Edition Lung
423 Cancer Stage Classification. *Chest* **151**, 193–203 (2017).
- 424 24. William D. Travis, Elisabeth Brambilla, H. Konrad Müller-Hermelink, & Curtis C. Harris.
425 *Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart*. (IARC
426 Press, Lyon, 2004).
- 427 25. Tamburro, D. *et al.* Analytical performance of a standardized kit for mass spectrometry-
428 based measurements of human glycosaminoglycans. *Journal of Chromatography B* **1177**,
429 122761 (2021).
- 430 26. Volpi, N., Galeotti, F., Yang, B. & Linhardt, R. J. Analysis of glycosaminoglycan-
431 derived, precolumn, 2-aminoacridone-labeled disaccharides with LC-fluorescence and
432 LC-MS detection. *Nat Protoc* **9**, 541–558 (2014).
- 433 27. Qvick, A. *et al.* Liquid biopsy as an option for predictive testing and prognosis in patients
434 with lung cancer. *Mol Med* **27**, 68 (2021).
- 435 28. Kruschke, J. K. Bayesian estimation supersedes the t test. *J Exp Psychol Gen* **142**, 573–
436 603 (2013).

- 437 29. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-
438 Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of
439 MCMC (with Discussion). *Bayesian Analysis* **16**, 667–718 (2021).
- 440 30. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal*
441 *of Statistical Software* **80**, 1–28 (2017).
- 442 31. Makowski, D., Ben-Shachar, M. S. & Lüdtke, D. bayestestR: Describing Effects and
443 their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of*
444 *Open Source Software* **4**, 1541 (2019).
- 445 32. Pál, D. *et al.* Compositional Analysis of Glycosaminoglycans in Different Lung Cancer
446 Types-A Pilot Study. *Int J Mol Sci* **24**, 7050 (2023).
- 447 33. Mattox, A. K. *et al.* The Origin of Highly Elevated Cell-Free DNA in Healthy Individuals
448 and Patients with Pancreatic, Colorectal, Lung, or Ovarian Cancer. *Cancer Discov* **13**,
449 2166–2179 (2023).
- 450 34. Bratulic, S. *et al.* Noninvasive detection of any-stage cancer using free
451 glycosaminoglycans. *Proceedings of the National Academy of Sciences* **119**,
452 e2115328119 (2022).
- 453 35. Wang, F. *et al.* Ultra-short cell-free DNA fragments enhance cancer early detection in a
454 multi-analyte blood test combining mutation, protein and fragmentomics. *Clin Chem Lab*
455 *Med* **62**, 168–177 (2024).
- 456 36. D’Ambrosi, S. *et al.* Combinatorial Blood Platelets-Derived circRNA and mRNA
457 Signature for Early-Stage Lung Cancer Detection. *Int J Mol Sci* **24**, 4881 (2023).
- 458 37. Bratulic, S. *et al.* Analysis of normal levels of free glycosaminoglycans in urine and
459 plasma in adults. *Journal of Biological Chemistry* **298**, (2022).
- 460 38. da Silva, M. N. R., Mendes, A., Martins, J. R. M., Tobias-Machado, M. & Pinhal, M. A.
461 da S. Prospective Evaluation of Chondroitin Sulfate, Heparan Sulfate and Hyaluronic
462 Acid in Prostate Cancer. *Int Braz J Urol* **44**, 1139–1146 (2018).

- 463 39. Saito, A. & Munakata, H. Analysis of plasma proteins that bind to glycosaminoglycans.
464 *Biochim Biophys Acta* **1770**, 241–246 (2007).
- 465 40. Prabhakar, V. & Sasisekharan, R. The biosynthesis and catabolism of
466 galactosaminoglycans. *Adv Pharmacol* **53**, 69–115 (2006).
- 467