



Published in final edited form as:

Anal Chem. 2013 March 5; 85(5): 2927–2936. doi:10.1021/ac3035579.

Multiplexed Surrogate Analysis of Glycotransferase Activity in Whole Biospecimens

Chad R. Borges^{1,*}, Douglas S. Rehder¹, and Paolo Boffetta²

¹Molecular Biomarkers Unit, The Biodesign Institute at Arizona State University, Tempe, AZ 85287

²Institute for Translational Epidemiology and Tisch Cancer Institute, Mount Sinai School of Medicine, New York, NY 10029

Abstract

Dysregulated glycotransferase enzymes in cancer cells produce aberrant glycans—some of which can help facilitate metastases. Within a cell, individual glycotransferases promiscuously help construct dozens of unique glycan structures, making it difficult to comprehensively track their activity in biospecimens—especially where they are absent or inactive. Here we describe an approach to deconstruct glycans in whole biospecimens then analytically pool together resulting monosaccharide-and-linkage-specific degradation products (“glycan nodes”) that directly represent the activities of specific glycotransferases. To implement this concept a reproducible, relative quantitation-based glycan methylation analysis methodology was developed that simultaneously captures information from N-, O-, and lipid linked glycans and is compatible with whole biofluids and homogenized tissues; in total over 30 different glycan nodes are detectable per GC-MS run. Numerous non-liver organ cancers are known to induce production of abnormally glycosylated serum proteins. Thus following analytical validation in blood plasma the technique was applied to a cohort of 59 lung cancer patient plasma samples and age/gender/smoking-status-matched non-neoplastic controls from the Lung Cancer in Central and Eastern Europe Study to gauge the clinical utility of the approach towards detection of lung cancer. Ten smoking-independent glycan node ratios were found that detect lung cancer with individual ROC c-statistics ranging from 0.76–0.88. Two glycan nodes provided novel evidence for altered ST6Gal-I and GnT-IV glycotransferase activities in lung cancer patients. In summary, a conceptually novel approach to the analysis of glycans in unfractionated human biospecimens has been developed that, upon clinical validation for specific applications, may provide diagnostic and/or predictive information in glycan-altering diseases.

Keywords

Glycans; glycomics; glycotransferase surrogates; whole biospecimens; lung cancer

INTRODUCTION

Glycans are complex, heterogeneous biological sugar polymers generally found attached to proteins or lipids and displayed on cell and macromolecule surfaces. The construction and display of abnormal glycan structures is an established hallmark of nearly every known type of tumor cell and appears to facilitate their ability to metastasize¹. In addition, there are

*Author to whom correspondence should be addressed: Chad R. Borges, Molecular Biomarkers Unit, The Biodesign Institute at Arizona State University, P.O. Box 876601, Tempe, AZ 85287. Tel 480-727-9929, fax 480-727-9464; chad.borges@asu.edu.

numerous types of cancer including ovarian²⁻³, prostate⁴⁻⁵, pancreatic⁶⁻⁷, liver⁸⁻⁹, multiple myeloma¹⁰, breast¹¹⁻¹², lung¹³⁻¹⁴, gastric¹⁵⁻¹⁶, thyroid¹⁷ and colorectal cancer¹⁸, as well as other inflammation-related diseases¹⁹⁻²⁰ that are able to induce aberrant glycosylation of abundant blood plasma proteins.

Glycans are created in the endoplasmic reticulum and golgi apparatus organelles by enzymes known as glycotransferases (GTs). Aberrant GT expression and/or activity is generally the immediate upstream cause of irregular glycan production¹. Unfortunately, however, the ability to directly track the activity of one or more GTs in human biospecimens is technically difficult and/or generally precluded in common clinical samples where GTs tend to lose activity *ex vivo* or are simply absent.

The natural complexity and structural heterogeneity of glycans comes in part from the fact that GTs build at glycan polymer branch-points and chain link sites in a non-template-driven, first-come-first-build manner—i.e., there are no biologically embedded templates or instruction sets that drive glycan construction in a precise, well-defined manner (such as is the case with DNA and proteins). Yet amidst this seemingly chaotic process, individual GTs generally exhibit strict donor, acceptor, and linkage specificity²¹, allowing for a moderate degree of consistency in routine glycan production.

When viewed across all protein and lipid substrates, the altered expression of a single GT can result in the production of a complex, heterogeneous mixture of n unique, abnormal whole-glycan structures rather than in uniformly increased expression of a single whole-glycan structure (Fig. 1). These heterogeneous mixtures of whole-glycan structures are difficult to fully characterize routinely—so existing cancer markers and novel candidate biomarkers that are based on intact glycan structure are generally based on one or a few particular aberrant glycan structures (out of n)—or perhaps a set of very closely related aberrant glycan structures that result in a unique antibody or lectin epitope.

With this background in mind, we developed the idea that monosaccharide-and-linkage-specific glycan polymer chain links and branch points (“glycan nodes”, as we refer to them), if broken down and quantified from the pool of all glycan structures in a biological sample may, in numerous cases, serve as direct, 1:1 molecular surrogates of aberrant GT activity—a complementary contrast to traditional glycomics approaches that focus on the analysis of whole, intact glycans that represent $1/n$: 1 molecular surrogates of GT activity (Fig. 1).

Below we describe the development and technical characteristics of a clinical sample-compatible protocol by which we have implemented this analytical concept. In the context of lung cancer, we provide an initial assessment of its utility as a methodology for routine measurement of novel glycan-based cancer markers.

EXPERIMENTAL SECTION

Materials

Heavy stable isotope-labeled D-Glucose (U-¹³C₆, 99%; 1,2,3,4,5,6-D7, 97–98%) was obtained from Cambridge Isotope Laboratories. N-acetyl-D-[UL-¹³C₆]glucosamine and L-[UL-¹³C₆]fucose were obtained from Omicron Biochemicals, Inc. 6'-Sialyl-N-acetyllactosamine and N-acetyllactosamine were purchased from Carbosynth (UK). Additional monosaccharide and glycan polymer standards for verification of partially methylated alditol acetate (PMAA) identities by GC-MS were obtained from Carbosynth, Sigma-Aldrich, V-Labs (a US subsidiary of Dextra UK), and The Scripps Research Institute / Consortium for Functional Glycomics. Repurified proteins were obtained from EMD Millipore (Human Serum Amyloid P), Sigma-Aldrich (Bovine Ribonuclease B), and

Athens Research & Technology (Human Vitamin D Binding Protein); pre-purified neutral glycosphingolipids were from Enzo Life Sciences. Sodium hydroxide beads (20–40 mesh) were purchased from Sigma-Aldrich. Spin columns (0.9 mL) equipped with plugs and polyethylene frits were purchased from the Pierce division of ThermoFisher Scientific (Cat. No. 69705). GC-MS autosampler vials and Teflon-lined pierceable caps were also obtained from ThermoFisher Scientific. GC consumables were acquired from Agilent; MS consumables were from Waters. All other solvents and chemicals were of the highest purity available and obtained from either ThermoFisher Scientific or Sigma-Aldrich.

Samples

A cohort of 59 blood plasma samples from lung cancer patients and age/gender/smoking-status matched controls that were enrolled in the Lung Cancer in Central and Eastern Europe (CEE) study were a gift from the International Agency for Research on Cancer biobank in Lyon, France. Additional serum samples from nominally healthy individuals, lung cancer patients and colorectal cancer patients were purchased from ProMedDx (Norton, MA). Serum samples from prostate cancer patients were purchased from the Cooperative Human Tissue Network (Vanderbilt, TN). Plasma samples from patients with normal glucose tolerance (NGT), impaired glucose tolerance (IGT), type 2 diabetes (T2D), and T2D with cardiovascular disease (CVD, defined as a history of heart attack or stroke and/or the presence of micro- or macroalbuminuria) were provided through an ongoing NIH-sponsored collaboration with Dr. Craig Stump and Dr. Hussein Yassine, endocrinologists at the University of Arizona. (Dr. Yassine is now at the University of Southern California.) Other biospecimens were purchased from Bioreclamation (Hicksville, NY), including a 300-mL plasma sample from an individual donor which was analyzed in every batch as a quality control sample.

Permethylation and Semi-purification of Whole Biofluid Glycans

Permethylation and subsequent clean-up procedures were adapted from the protocol of Goetz et al²², which was designed to permethylate and release O-linked glycans from pre-isolated glycoproteins. Nine microliters of a whole biofluid sample (e.g., blood plasma, serum, seminal fluid, homogenized tissue etc.) was added to a 1.5 mL polypropylene test tube. To this was added 1 μ L of an internal tracer stock solution containing 10 mM each of D-Glucose (U-¹³C₆, 99%; 1,2,3,4,5,6,6-D7, 97–98%), N-acetyl-D-[UL-¹³C₆]glucosamine and L-[UL-¹³C₆]fucose. (As explained below these internal standards are useful for qualitative verification of proper sample processing, but for purposes of relative quantification the glycan nodes within a sample are best normalized to themselves.) To the 10- μ L sample-plus-internal-standard mixture was added 270 μ L of dimethylsulfoxide (DMSO) and 105 μ L of iodomethane. This solution was mixed thoroughly and placed onto a plugged 1-mL spin column containing approximately 0.7 g sodium hydroxide beads that had been preconditioned with acetonitrile followed by two rinses with DMSO. Samples were allowed to sit for 10–12 minutes with occasional stirring. Samples were then unplugged and spun in a microcentrifuge for 30 s at 4,000 rpm (800 g) to retrieve the glycan-containing liquid. Samples were then transferred to a silanized 13 \times 100 glass test tube. Three hundred microliters of acetonitrile was then added to the spin column to wash off all of the permethylated glycan. Spin columns were then centrifuged at 10,000 rpm (5,000 g) for 30 s to collect the acetonitrile which was pooled with the rest of the sample. To the liquid sample was added 3.5 mL of 0.5 M NaCl followed by 1.2 mL of chloroform. Liquid/liquid extraction was performed 3 times, saving the chloroform layers, which were dried under a gentle stream of nitrogen.

Glycan Methylation Analysis

The following procedure was adapted from Heiss et al²³:

Trifluoroacetic acid (TFA) hydrolysis—Three hundred twenty five microliters of 2 M TFA was added to each sample which was then capped tightly and heated at 121 °C for 2 hrs. TFA was then removed by heat-assisted evaporation under a gentle stream of nitrogen.

Reduction of sugar aldehydes—A fresh 10 mg/ml solution of sodium borohydride in freshly prepared 1 M ammonium hydroxide was prepared and added (475 µL) to each test tube, mixed thoroughly and allowed to react for 1 hour at room temperature. Residual borate was removed by adding 5 drops of methanol to each sample, drying under nitrogen, then adding 125 µL of 9:1 (v/v) MeOH:Acetic Acid and drying again under nitrogen. Samples were then dried for about 30 minutes in a vacuum desiccator before proceeding.

Acetylation of nascent hydroxyl groups—Two hundred and fifty microliters of freshly made water-saturated acetic anhydride (16:234 v/v water:acetic anhydride) was added to each sample, which was mixed thoroughly to dissolve as much of the sample residue as possible. Next, 230 µL of concentrated TFA was added to each sample, which was then mixed, capped and incubated at 50 °C for 10 minutes.

Final Clean-up—Two milliliters of dichloromethane was added to each sample along with 2 milliliters of water. Liquid/liquid extraction carried out twice with water. The final organic layer was then dried in a silanized autosampler vial under nitrogen and reconstituted in 8 drops of acetone, mixed, capped and placed on the GC-MS autosampler rack.

Overall sample throughput is limited by the time required for sample preparation. One analyst can reasonably process ~ 60–75 samples per week. We anticipate that automated sample processing robotics will be able to significantly reduce this bottleneck in throughput.

Gas Chromatography-Mass Spectrometry

GC/MS was carried out on an Agilent A7890 gas chromatograph (equipped with a CTC PAL autosampler) coupled to a Waters GCT (Time-of-Flight) mass spectrometer.

One microliter was injected in split mode onto an Agilent split-mode liner (Cat. No. 5183-4647) containing a small plug of silanized glass wool, maintained at 280 °C. All injections were made in duplicate: once at a split ratio of 50 and once at a split ratio of 75. Using helium as the carrier gas (0.8 mL/min, constant flow mode) samples were chromatographed over a 30 m DB-5ms GC column. The oven was initially held at 165 °C for 0.5 minutes followed by ramping at 10 °C/min to 265 °C then immediately ramping at 30 °C/min to 325 °C and holding for 3 minutes (15.5 minutes of total run time). The transfer line was maintained at 250 °C. Sample components eluting from the GC column were subjected to electron ionization (70 eV, 250 °C) and analyzed from m/z 40 to 800 with a “scan cycle” time of 0.2 s. The mass spectrometer was tuned and calibrated (to within 10 ppm mass accuracy) daily using perfluorotributylamine.

Data Analysis

Initial identification of PMAAs was made through the analysis of glycan standards and verified through comparison with the online electron ionization mass spectral library of PMAAs at the University of Georgia’s Complex Carbohydrate Research Center: <http://www.crc.edu/databases/index.php#>

The top most abundant and/or diagnostic fragment ions for each glycan node in blood plasma/serum were summed (using a 0.15 Da extracted ion chromatogram mass window) for quantification (Supporting Information Table S1). Quantification was carried out by integration of summed extracted ion chromatogram peak areas in automated fashion using

QuanLynx Software. Integrated peaks were manually verified then exported to a spreadsheet for further calculation.

All statistical analyses including generation of ROC curves were carried out using XLSTAT Version 2012.3.01. All t-tests for significant differences between group means were two-sided and pre-evaluated for variance scedasticity. No weighting was employed during ANOVA or ROC calculations.

RESULTS AND DISCUSSION

Strategy and Initial Development

For decades now, glycan methylation analysis (Fig. 2) has been employed to collect monosaccharide-specific linkage information from pre-isolated glycans. Given the fact that numerous monosaccharide-specific linkage patterns (i.e., glycan nodes) are created by one or just a few GTs (Supporting Information Table S2), our goal was to enable glycan methylation analysis as a biomarker development tool by adapting it for routine use with common human biological specimens—minimizing or eliminating sample pre-fractionation steps.

Solid-phase sodium hydroxide-based permethylation procedures were first developed by Cicanu and co-workers^{24–26} and later refined into online and spin-column based approaches by Kang et al^{27–28}. A spin-column based procedure reported by Goetz, Novotny and Mechref in 2009²² reported the specific chemical release of O-glycans from intact proteins and for us represented a promising front-end preparatory step. With little modification we found that this method could not only release O-linked protein glycans but, when coupled with a TFA-based methylation analysis protocol²³, resulted in the release and detection of partially methylated alditol acetates (PMAAs) from N-linked protein glycans as well as glycolipids (Fig. 3). As expected all forms of hexose (plus xylose) and N-acetylhexosamine (HexNAc) residues were detectable. Sulfated, hexuronic and sialic acid residues were only detected *indirectly* vis-à-vis their linkage positions. Reducing-end monosaccharides in N-linked glycans appeared in the final analysis, unaffected by their unique nitrogen atom linkage. This was evidenced by the routine detection of the PMAA corresponding to 4,6-linked GlcNAc (4,6-GlcNAc) in blood plasma samples (Fig. 1): Based on the database research conducted to create Supporting Information Table S2, there is only one GT capable of producing this glycan node—alpha-(1,6)-fucosyltransferase—which exclusively catalyzes the addition of a fucose residue to the 6-position of the reducing end 4-linked GlcNAc residue in N-glycans.

The suitability of this approach for direct application with 10-NL volumes of whole biofluids and homogenized tissue samples was assessed and, in every biomatrix tested to date, proved qualitatively compatible (Figs. 1 and 4). These biomatrix compatibility findings opened up the technique to a wide variety of potential clinical applications. In practical terms, however, blood plasma represented the most readily available biomatrix for assessing the potential clinical utility of the technology.

Evaluation in Lung Cancer

Following an initial evaluation of reproducibility in blood plasma (described below), we applied the new analytical technology to a cross-sectional pilot study of 59 archived blood plasma samples from patients enrolled in the Lung Cancer in Central and Eastern Europe (CEE) study.²⁹ Summary information on gender, age, and smoking history is shown in Table 1. Additional detailed information on the patients enrolled in this study can be found online (Supporting Information Tables S3-S4). In most cases samples from lung cancer patients were taken within a few days of initial diagnosis. Controls in this study were

matched to the lung cancer patients by age, gender, and smoking status and were enrolled upon visiting participating clinics for non-neoplastic conditions unrelated to tobacco smoking.

Randomized samples were analyzed blind in six separate batches. Despite the addition of heavy, stable-isotope labeled monosaccharides to each sample as internal standards we found that, in general, the ratios of endogenous glycan nodes to each other (GNRs) tended to provide greater analytical precision than the ratios of individual glycan nodes to stable-isotope labeled internal standards (iGNs) (Table 2 and Supporting Information Table S5, described in additional detail below). In the CEE cohort eight iGNs and 29 GNRs were found to be significantly different ($p < 0.005$) in the lung cancer cases vs. controls. The top two performing GNRs had receiver operating characteristic (ROC) *c*-statistics in the 0.8 – 0.9 range (Fig. 5). The top 12 performing GNRs had ROC *c*-statistics > 0.75 —better than any single iGN (Table 2 and Supporting Information Table S5).

To evaluate if these GNRs might be mere indicators of smoking status the ROC curve analysis was repeated for smokers only and on the basis of smokers (including current and former) vs. non-smokers (i.e., never-smokers / no smoking history), regardless of cancer status (Table 2). ROC curve analysis for smokers-only demonstrated negligible difference compared to when non-smokers were included in the analysis (Table 2). ROC curve analysis for smokers vs. nonsmokers (regardless of cancer status) demonstrated an across-the-board loss of diagnostic power for 10 of the top 12 GNRs (Table 2)—indicating that these markers are linked to the presence of cancer and not smoking history. Interestingly, the two GNR ROC *c*-statistics that remained the same or increased in this comparison had the same common denominator (3-GalNAc). The biological relationship of this glycan node to smoking, if any, is not yet clear.

Analytical Validation in Blood Plasma

Analytical validation of the approach in blood plasma was undertaken with the goals of determining reproducibility (intra- and inter-day precision), sample stability, consistency of results in serum and four different types of plasma, autosampler stability, and analytical sensitivity and linearity of response. Due to space constraints and the need to avoid presentation of copious amount of superfluous data, the analytical validation parameters described below are largely contextualized to the top 12 performing GNRs in lung cancer (Table 2).

Precision / Reproducibility—Despite the addition of heavy, stable-isotope labeled monosaccharides to each plasma sample as internal standards we found that, based on the analysis of six aliquots of the same plasma sample per day on three different days, the ratios of individual glycan nodes to stable-isotope labeled internal standards (iGNs) were not highly reproducible (Supporting Information Table S5) and that the ratios of endogenous glycan nodes to each other (GNRs) tended to provide greater analytical precision. Since approximately 20 iGNs were routinely detected in plasma samples, this meant over that 200 GNRs were available for assessment of reproducibility. Of these GNRs over 80 had both intra and inter-ssay reproducibility of less than 15%. The analytical precision of the top 12 performing GNRs in lung cancer are reported in Table 2.

Sample Stability—GNR stability in blood plasma samples was assessed by creating twelve aliquots of a single sample then placing six back into the $-80\text{ }^{\circ}\text{C}$ freezer and leaving the remaining six at room temperature overnight. None of the top six, but two of the top 12 lung cancer GNRs demonstrated statistically significant difference between the batches (Supporting Information Table S6). These apparent differences, however, were subtle and

may have been due to abnormally tight intra-assay precision as the overall differences between the batches were less than 12%.

Effect of Blood Collection Type—Matched sets of blood serum, 3.8% sodium citrate, Na₂EDTA, K₂EDTA, and K₃EDTA plasma samples were collected from 22 healthy volunteers. Glycan nodes were then analyzed in the resulting 110 samples. Analysis of the results for the top 12 lung cancer GNRs by repeated measures ANOVA followed by the Ryan-Einot-Gabriel-Welsch (REGW) multiple comparison test demonstrated no significant differences between the blood collection types (Supporting Information Table S7).

Autosampler Stability—Autosampler stability was assessed over the time span required to inject 48 samples (approximately 14.5 hrs)—the largest batch size employed during analysis of samples reported in this paper. Four data points from the same sample were acquired three times—at the beginning, middle and end of this time period. Passing stability was designated to be within 10% of the average of the first two data points. Autosampler stability passed for the top 12 lung cancer GNRs except the three with HexNAc nodes as the denominator (Supporting Fig. S1)—which were of minimal interest since 2 of these 3 appear to be more diagnostic of smoking rather than lung cancer and the third is not in the top 6 GNRs (Table 2).

Analytical Sensitivity and Linearity of Response—IUPAC defines analytical sensitivity as the ability of an analytical procedure to produce a change in signal for a defined change in analyte quantity.³⁰ They add that in most cases, this parameter can be observed as the slope of a calibration curve. The fact that strong, statistically significant differences were detected in numerous iGNs and GNRs between the CEE lung cancer cases and controls (Table 2 and Supporting Information Table S5) suggested that analytical sensitivity was more than adequate to impart the technique with potential clinical applicability—but the expense and limited availability of glycan polymer standards precluded a formal assessment of the analytical sensitivity for all 12 of the top lung cancer GNRs. The instrument response ratio of t-Gal/6-Gal vs. the actual molar ratio of t-Gal/6-Gal was assessed, however, through the use of Nacetyllactosamine (Gal1-4GalNAc) and 6-Sialyl-N-acetyllactosamine (N-euNAc2-6Gal1-4GalNAc) standards. These were mixed together in ratios spanning the physiologically observed range in blood plasma and such that overall signal intensities approximated those from blood plasma samples. Following analysis on two different days, the resulting data were employed to construct a standard curve (Supporting Information Fig. S2). The standard curve was linear ($R^2 = 0.993$) and had a slope of 1.31 indicating a greater than 1:1 change in instrument response per change in actual glycan node molar ratio.

Other Analytical Validation Considerations—The goal of this study was to evaluate *changes* in the relative abundance of readily detectable glycan nodes as potential clinical markers—a distinct and separate goal from quantifying low abundance glycan nodes. As such, unlike most conventional assays, *raw sensitivity (i.e., limits of detection and limits of quantification)* was not a parameter of significant concern because, for each biomatrix, a particular set of glycan nodes was present in every sample at readily detectable levels. For example, though not visible with the particular extracted ion chromatograms from plasma shown in Fig. 1, there were over 18 individual glycan nodes with signal/noise (S/N) ratios greater than 10 in every plasma sample of hundreds of individual samples tested to date. Thus, in the absence of a specific need to detect low abundance glycan node(s), raw detection limits were of little practical importance or value relative to analytical sensitivity. In the future, if raw detection limits are required for a particular application, they will need to be investigated on a biomatrix-specific basis.

Since this is a technique for the relative quantification of the constituent components of heterogeneous biological polymers, *accuracy* cannot be defined by any single molecular standard. But this does not rule out clinical utility or applicability: To achieve these things without a definition of absolute accuracy good reproducibility/precision of iGN/GNR measurement will be critical (as documented above), but it will eventually have to be coupled with a mechanism to facilitate inter-laboratory transferability—for example, through establishment of a “gold standard” sample that can be shared across laboratories or through instrument-specific calibration with pre-defined standard curves (such as that in Supporting Information Fig. S2). Since the latter will have to be based on particular chemical standards, they still will not be able to define accuracy in an absolute sense and, at best, will only ever be considered to provide results that are *approximately* accurate according to strict definition—but this is essentially irrelevant when it comes to practical application.

Disease Specificity

To further evaluate the specificity of the top-performing blood plasma GNRs for lung cancer two additional sets of samples were analyzed for comparison: The first set consisted of biobank-purchased serum samples from 80 healthy individuals, an additional 16 lung cancer patients, 10 colorectal cancer and 59 prostate cancer patients. The second set consisted of a cross-section of patients from a University of Arizona diabetes study who had undergone an oral glucose tolerance test and ranged from healthy (normal glucose tolerance, NGT, n = 18), to pre-diabetic (impaired glucose tolerance, IGT, n = 12), to stark type 2 diabetes (T2D, n = 32), to T2D with cardiovascular complications (T2D w/ CVD n = 26). Analysis of the top 6 lung cancer GNRs (Table 2) in these samples by ANOVA followed by the REGW multiple comparison test demonstrated a general grouping of lung cancer and colorectal cancer separate from the other cohorts, with T2D aligning more closely with lung and colorectal cancer than did prostate cancer (Supporting Information Fig. S3). These cross-disease comparisons suggested that the GNR markers in blood were not just fluctuating with inflammation and possessed at least a limited degree of specificity for certain types of cancer.

Comparison of GNRs in biobank-purchased serum samples from cancer patients with those from nominally healthy individuals revealed general consistency of GNR behavior in the two lung cancer cohorts (Table 2 and Supporting Information Table S8). In addition, there was partial GNR profile overlap between lung and colorectal cancer but not prostate cancer (Table 2 and Supporting Information Tables S8-S10). Based on increases in ROC c-statistics, GNRs were better at distinguishing lung cancer patients from fully healthy individuals than from well-match controls (Table 2 and Supporting Information Table S8)—an unsurprising but useful observation to note when it comes to the design of biomarker studies. GNRs in serum did not appear to be particularly diagnostic in prostate cancer (Supporting Information Table S10).

Biological Implications

Glycan nodes observed via this technique are necessarily derived from the most abundant glycan source in the sample under consideration. In blood plasma, roughly half of the glycoproteins are immunoglobulins / complement protein and the other half are liver glycoproteins. Unless near-milligram-per-milliliter plasma concentrations are reached, cancer glycoprotein shedding is unlikely to contribute more than an unobservable fraction to total plasma glycoprotein content. This means that cancer-induced alterations to the humoral immune system and/or the liver are most likely responsible for the alterations in plasma glycan nodes observed here. Though the mechanisms behind such phenomena (e.g., those discovered by Narisada et al³¹ and Kitazume et al³²) are varied and the phenomena

themselves are only partially understood, they are by no means unknown,^{2–18}—even in non-neoplastic diseases.^{19–20}

Glycan source notwithstanding, aberrant activity of at least two different GTs in lung and colorectal cancer patients was detected (Table 2 and Supporting Information Table S5 and Supporting Information Tables S8–S10). A table of GTs responsible for producing the glycan nodes observed in humans is provided online (Supporting Information Table S2): 1) ST6Gal-I: Six of the top 12-performing GNRs in lung cancer (Table 2) and five of the top seven GNRs in colorectal cancer (Supporting Information Table S9) involved t-Gal and/or 6-Gal. When their behavior was considered in cancer cases compared to controls these GNRs showed fluctuations consistent with minor changes in t-Gal ($p = 0.04$) and significant increases in 6-Gal ($p = 8.6 \times 10^{-4}$; ROC c-statistic = 0.737 ± 0.064 ; Supporting Information Table S5). Together these data indicated increased β -galactoside: α 2-6sialyltransferase (ST6Gal-I) enzyme activity—a phenomenon for which there is evidence in cancer cells from numerous other carcinomas including those of the colon/rectum,^{33–34} breast,³⁵ brain (non-neuroectodermal epithelial-like tumors),³⁶ cervix,³⁷ and liver (transgenic mouse model of hepatocellular carcinoma),³⁸ as well as in choriocarcinoma (cell lines)³⁹ and acute myeloid leukemia.⁴⁰ 2) GnT-IV: Five of the top 12 lung cancer GNRs (Table 2) and two of the top seven colorectal cancer GNRs (Supporting Information Table S9) provided evidence for elevated quantities of 2,4-Man. In addition, the 2,4-Man iGN in CEE lung cancer patients was significantly higher than in the controls ($p = 3.4 \times 10^{-4}$; ROC c-statistic = 0.747 ± 0.063 ; Supporting Information Table S5). Increased 2,4-Man is mediated through increased UDP-N-acetylglucosamine: α -1,3-D-mannosidase β 1,4-N-acetylglucosaminyltransferase IV (GnT-IV) activity. In general this enzyme has been documented as overactive during oncogenesis and differentiation.⁴¹ Evidence for its overexpression has been found in cancer cells from colorectal carcinoma,⁴² choriocarcinoma,⁴³ hepatocellular carcinoma,⁴⁴ and pancreatic cancer (GnT-IVb form).⁴⁵

To our knowledge this is the first report of altered activity for either ST6Gal-I or GnT-IV in lung cancer patients; in our opinion, the fact that these changes may not be derived directly from tumor cells themselves makes the findings intriguing—particularly with regard to potential therapeutic implications embedded in the underlying mechanism(s). Based on the CEE cohort, a total of 12 iGNs were significantly increased in lung cancer patient plasma, providing evidence for activity changes in numerous other GTs as well, including fucosyltransferases (vis-à-vis increased t-Fuc and 3,4-GlcNAc) and GnT-V (vis-à-vis increased 2,6-Man) (Supporting Information Tables S2 and S5).

Four of the seven increasing GNRs in lung cancer (Table 2) contained 3,4,6-Man as their denominator, providing suggestive evidence for decreased β 1,4-N-acetylglucosaminyltransferase III (GnT-III) activity—the GT responsible for adding “bisecting GlcNAc” to N-linked glycans. However, the change in the 3,4,6-Man iGN was not statistically significant. Notably, the 3,6-Man iGN increased in CEE lung cancer patients relative to controls ($p = 1.4 \times 10^{-3}$; Supporting Information Table S5)—suggesting an overall increase in N-glycans and a relative inability of GnT-III to keep pace.

CONCLUSIONS

There is an urgent need for blood-borne markers of risk and progression in lung cancer as well as other types of cancer (and glycan-affective disorders) to which this analytical approach will likely be applicable. Based on current knowledge of human GTs, the glycan nodes 6-Gal, 2,4-Man, 2,6-Man, and 3,4,6-Man represent 1:1 (or nearly 1:1) molecular surrogates for ST6Gal-I, GnT-IV, GnT-V and GnT-III respectively—and there are other glycan nodes that hold this same relationship with their respective GTs (Supporting

Information Table S2). Additional glycan nodes, such as 3-Gal, may represent the activity of multiple GTs but they are still potentially modulated by the aberrant expression of just one of the GTs that leads to their existence.

By condensing and pooling together into a single analytical signal the inherent molecular heterogeneity introduced by aberrantly expressed GTs (Fig. 1)—and doing so for multiple GTs simultaneously from 10 μ L of whole, unprocessed biofluid without the use of enzyme or antibody reagents—this analytical approach represents a promising means by which to access glycans as disease markers. Its utility is expected to improve further once it can be applied to hundreds of well-characterized patient samples for which outcome information is available, then coupled to multivariate modeling algorithms to create disease-specific prognostic biosignatures. Finally, we note that cancer-induced aberrant glycans in plasma/serum may be diluted by significant quantities of normal glycans; as such, going forward we feel that application of this technology to biofluids or tissues obtained directly from putatively cancerous organs (e.g., as demonstrated in Fig. 4) may represent a more powerful use of this technology to address specific medical needs for better cancer markers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by a Flinn Foundation award and, in part, by NIH grants R01DK082542 and R24DK083948, as well as grant number GM62116 to the Consortium for Functional Glycomics. We thank Dr. Mark Holl and Dr. Mia Hashibe for helpful discussions. We also thank the national PIs of the CEE study for allowing the use of the samples: Drs. D. Zaridze, Moscow, Russia; J. Lissowska, Warsaw, Poland; N. Szeszenia-Dabrowska, Lodz, Poland; E. Fabianova, Banska-Bystrica, Slovakia; P. Rudnai, Budapest, Hungary; D. Mates, Bucharest, Romania; V. Bencko, Prague, Czech Republic; L. Foretova, Brno, Czech Republic; V. Janout, Olomouc, Czech Republic.

Abbreviations

GTs	Glycotransferase enzymes
iGN	individual glycan node
GNR	glycan node ratio
GC-MS	gas chromatograph-mass spectrometry
CEE	Lung Cancer in Central and Eastern Europe Study
ROC	receiver operating characteristic
NGT	normal glucose tolerance
IGT	impaired glucose tolerance
T2D	type 2 diabetes
CVD	cardiovascular disease

REFERENCES

1. Varki, A.; Kannagi, R.; Toole, B. *Essentials of Glycobiology*. 2nd edn.. Varki, A.; Cummings, RD.; Esko, JD.; Freeze, HH.; Stanley, P.; Bertozzi, CR.; Hart, GW.; Etzler, ME., editors. New York: Cold Spring Harbor Laboratory Press; 2009.
2. Gercel-Taylor C, Bazzett LB, Taylor DD. *Gynecol Oncol*. 2001; 81:71–76. [PubMed: 11277653]

3. An HJ, Miyamoto S, Lancaster KS, Kirmiz C, Li B, Lam KS, Leiserowitz GS, Lebrilla CB. *J Proteome Res.* 2006; 5:1626–1635. [PubMed: 16823970]
4. Kanoh Y, Mashiko T, Danbara M, Takayama Y, Ohtani S, Egawa S, Baba S, Akahoshi T. *Anticancer Res.* 2004; 24:3135–3139. [PubMed: 15510601]
5. Kyselova Z, Mechref Y, Al Bataineh MM, Dobrolecki LE, Hickey RJ, Vinson J, Sweeney CJ, Novotny MV. *J Proteome Res.* 2007; 6:1822–1832. [PubMed: 17432893]
6. Okuyama N, Ide Y, Nakano M, Nakagawa T, Yamanaka K, Moriwaki K, Murata K, Ohigashi H, Yokoyama S, Eguchi H, Ishikawa O, Ito T, Kato M, Kasahara A, Kawano S, Gu J, Taniguchi N, Miyoshi E. *Int J Cancer.* 2006; 118:2803–2808. [PubMed: 16385567]
7. Zhao J, Patwa TH, Qiu W, Shedden K, Hinderer R, Misek DE, Anderson MA, Simeone DM, Lubman DM. *J Proteome Res.* 2007; 6:1864–1874. [PubMed: 17428079]
8. Comunale MA, Lowman M, Long RE, Krakover J, Philip R, Seeholzer S, Evans AA, Hann HW, Block TM, Mehta AS. *J Proteome Res.* 2006; 5:308–315. [PubMed: 16457596]
9. Goldman R, Resson HW, Varghese RS, Goldman L, Bascug G, Loffredo CA, Abdel-Hamid M, Gouda I, Ezzat S, Kyselova Z, Mechref Y, Novotny MV. *Clinical Cancer Research.* 2009; 15:1808–1813. [PubMed: 19223512]
10. Aurer I, Lauc G, Dumic J, Rendic D, Maticic D, Milos M, Heffer-Lauc M, Flogel M, Labar B. *Coll Antropol.* 2007; 31:247–251. [PubMed: 17598409]
11. Abd Hamid UM, Royle L, Saldova R, Radcliffe CM, Harvey DJ, Storr SJ, Pardo M, Antrobus R, Chapman CJ, Zitzmann N, Robertson JF, Dwek RA, Rudd PM. *Glycobiology.* 2008; 18:1105–1118. [PubMed: 18818422]
12. Kyselova Z, Mechref Y, Kang P, Goetz JA, Dobrolecki LE, Sledge GW, Schnaper L, Hickey RJ, Malkas LH, Novotny MV. *Clinical Chemistry.* 2008; 54:1166–1175. [PubMed: 18487288]
13. Hongsachart P, Huang-Liu R, Sinchaikul S, Pan FM, Phutrakul S, Chuang YM, Yu CJ, Chen ST. *Electrophoresis.* 2009; 30:1206–1220. [PubMed: 19294700]
14. Arnold JN, Saldova R, Galligan MC, Murphy TB, Mimura-Kimura Y, Telford JE, Godwin AK, Rudd PM. *J Proteome Res.* 2011; 10:1755–1764. [PubMed: 21214223]
15. Bones J, Mittermayr S, O'Donoghue N, Guttman A, Rudd PM. *Anal Chem.* 2010; 82:10208–10215. [PubMed: 21073175]
16. Kodar K, Stadlmann J, Klaamas K, Sergeev B, Kurtenkov O. *Glycoconj J.* 2012; 29:57–66. [PubMed: 22179780]
17. Chen G, Wang Y, Qiu L, Qin X, Liu H, Wang X, Wang Y, Song G, Li F, Guo Y, Li F, Guo S, Li Z. *J Proteomics.* 2012; 75:2824–2834. [PubMed: 22365975]
18. Takeda Y, Shinzaki S, Okudo K, Moriwaki K, Murata K, Miyoshi E. *Cancer.* 2012; 118:3036–3043. [PubMed: 22006099]
19. Parekh RB, Dwek RA, Sutton BJ, Fernandes DL, Leung A, Stanworth D, Rademacher TW, Mizuochi T, Taniguchi T, Matsuta K, et al. *Nature.* 1985; 316:452–457. [PubMed: 3927174]
20. Mehta AS, Long RE, Comunale MA, Wang M, Rodemich L, Krakover J, Philip R, Marrero JA, Dwek RA, Block TM. *J Virol.* 2008; 82:1259–1270. [PubMed: 18045939]
21. Rini, J.; Esko, J.; Varki, A. *Essentials of Glycobiology.* 2nd edn. Varki, A.; Cummings, RD.; Esko, JD.; Freeze, HH.; Stanley, P.; Bertozzi, CR.; Hart, GW.; Etzler, ME., editors. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.
22. Goetz JA, Novotny MV, Mechref Y. *Anal Chem.* 2009; 81:9546–9552. [PubMed: 19874002]
23. Heiss C, Klutts JS, Wang Z, Doering TL, Azadi P. *Carbohydrate research.* 2009; 344:915–920. [PubMed: 19345342]
24. Ciucanu I, Kerek F. *Carbohydr Res.* 1984; 131:209–217.
25. Ciucanu I, Costello CE. *Journal of the American Chemical Society.* 2003; 125:16213–16219. [PubMed: 14692762]
26. Ciucanu I, Caprita R. *Anal Chim Acta.* 2007; 585:81–85. [PubMed: 17386650]
27. Kang P, Mechref Y, Klouckova I, Novotny MV. *Rapid Communications in Mass Spectrometry.* 2005; 19:3421–3428. [PubMed: 16252310]
28. Kang P, Mechref Y, Novotny MV. *Rapid Communications in Mass Spectrometry.* 2008; 22:721–734. [PubMed: 18265433]

29. Brennan P, Crispo A, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, Lissowska J, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Fletcher T, Boffetta P. *Am J Epidemiol*. 2006; 164:1233–1241. [PubMed: 17032696]
30. McNaught, AD.; Wilkinson, A. IUPAC. Compendium of Chemical Terminology. XML on-line corrected version: <http://goldbook.iupac.org> (2006-) created by M. Nic J. Jirat, B. Kosata; updates compiled by A. Jenkins. doi:10.1351/goldbook. Last update: 2012-08-19; version: 2.3.2. 2nd (the "Gold Book") ed.. Oxford: Blackwell Scientific Publications; 1997.
31. Narisada M, Kawamoto S, Kuwamoto K, Moriwaki K, Nakagawa T, Matsumoto H, Asahi M, Koyama N, Miyoshi E. *Biochemical and Biophysical Research Communications*. 2008; 377:792–796. [PubMed: 18951869]
32. Kitazume S, Oka R, Ogawa K, Futakawa S, Hagiwara Y, Takikawa H, Kato M, Kasahara A, Miyoshi E, Taniguchi N, Hashimoto Y. *Glycobiology*. 2009; 19:479–487. [PubMed: 19150807]
33. Vierbuchen MJ, Fruechtnicht W, Brackrock S, Krause KT, Zienkiewicz TJ. *Cancer*. 1995; 76:727–735. [PubMed: 8625173]
34. Dall'Olio F, Malagolini N, di Stefano G, Minni F, Marrano D, Serafini-Cessi F. *Int J Cancer*. 1989; 44:434–439. [PubMed: 2476402]
35. Recchi MA, Hebbar M, Hornez L, Harduin-Lepers A, Peyrat JP, Delannoy P. *Cancer Res*. 1998; 58:4066–4070. [PubMed: 9751611]
36. Kaneko Y, Yamamoto H, Kersey DS, Colley KJ, Leestma JE, Moskal JR. *Acta Neuropathol*. 1996; 91:284–292. [PubMed: 8834541]
37. Wang PH, Li YF, Juang CM, Lee YR, Chao HT, Tsai YC, Yuan CC. *Gynecol Oncol*. 2001; 83:121–127. [PubMed: 11585423]
38. Pousset D, Piller V, Bureaud N, Monsigny M, Piller F. *Cancer Res*. 1997; 57:4249–4256. [PubMed: 9331085]
39. Fukushima K, Hara-Kuge S, Seko A, Ikehara Y, Yamashita K. *Cancer Res*. 1998; 58:4301–4306. [PubMed: 9766657]
40. Skacel PO, Edwards AJ, Harrison CT, Watkins WM. *Blood*. 1991; 78:1452–1460. [PubMed: 1679356]
41. Taniguchi N, Korekane H. *BMB Rep*. 2011; 44:772–781. [PubMed: 22189679]
42. D'Arrigo A, Belluco C, Ambrosi A, Digito M, Esposito G, Bertola A, Fabris M, Nofrate V, Mammanno E, Leon A, Nitti D, Lise M. *Int J Cancer*. 2005; 115:256–262. [PubMed: 15688387]
43. Endo T, Nishimura R, Kawano T, Mochizuki M, Kobata A. *Cancer Res*. 1987; 47:5242–5245. [PubMed: 2957051]
44. Yamashita K, Totani K, Iwaki Y, Takamisawa I, Tateishi N, Higashi T, Sakamoto Y, Kobata A. *J Biochem*. 1989; 105:728–735. [PubMed: 2568988]
45. Ide Y, Miyoshi E, Nakagawa T, Gu J, Tanemura M, Nishida T, Ito T, Yamamoto H, Kozutsumi Y, Taniguchi N. *Biochem Biophys Res Commun*. 2006; 341:478–482. [PubMed: 16434023]
46. Tennent GA, Pepys MB. *Biochemical Society Transactions*. 1994; 22:74–79. [PubMed: 8206305]
47. Kiernan UA, Nedelkov D, Tubbs KA, Niederkofler EE, Nelson RW. *Proteomics*. 2004; 4:1825–1829. [PubMed: 15174148]
48. Uniprot Consortium. *Nucleic Acids Research*. 2012; 40:D71–D75. [PubMed: 22102590]
49. Prien JM, Ashline DJ, Lapadula AJ, Zhang H, Reinhold VN. *J Am Soc Mass Spectrom*. 2009; 20:539–556. [PubMed: 19181540]
50. Borges CR, Jarvis JW, Oran PE, Nelson RW. *J Proteome Res*. 2008; 7:4143–4153. [PubMed: 18686987]
51. Borges CR, Jarvis JW, Oran PE, Rogers SP, Nelson RW. *J Biomol Tech*. 2008; 19:167–176. [PubMed: 19137103]

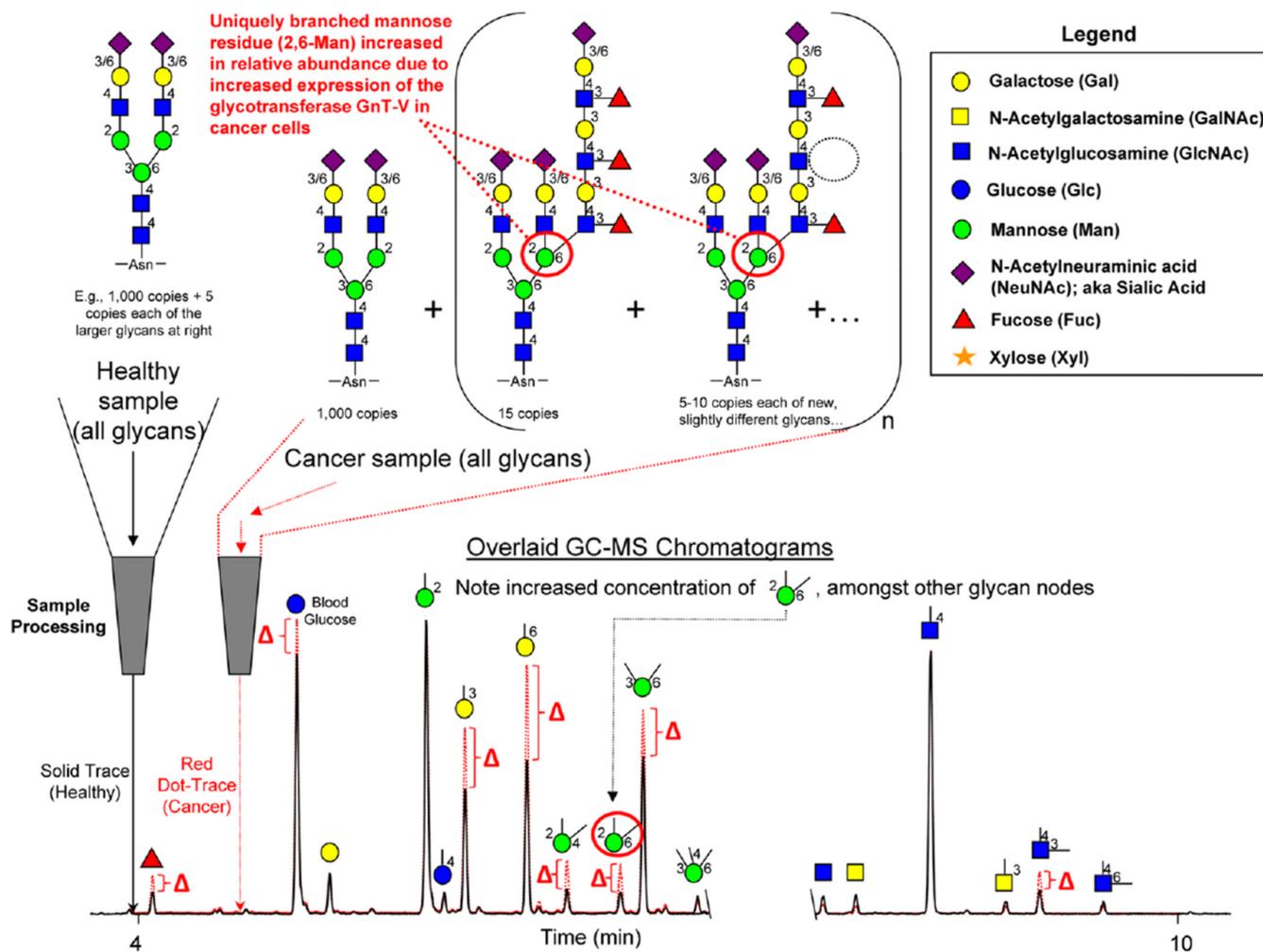


Figure 1.

Conceptual overview of the analytical concept: An upregulated GT (e.g., GnT-V) causes an increase in the quantity of a specific, uniquely linked glycan monosaccharide residue (a 2,6-linked Mannose “node” in this example)—which, through the subsequent action of other GTs, can lead to formation of a mixture of heterogeneous whole-glycan structures at low copy number each—all of which can be difficult to detect and quantify in routine fashion. Analytically pooling together the “glycan nodes” from amongst all the aberrant glycan structures provides a more direct surrogate measurement of GnT-V activity than any single intact glycan. Simultaneous measurement of N-, O-, and lipid linked “glycan nodes” in whole biospecimens as described here represents a conceptually novel means by which to detect and monitor glycan-affective diseases such as cancer. Actual extracted ion chromatograms from 10-microliter blood plasma samples shown. Numbers adjacent to monosaccharide residues in glycan structures indicate the position at which the higher residue is linked to the lower residue. If no linkage positions are indicated in the chromatogram annotation the residue is either in the terminal position or free in solution (e.g. glucose). All residues except sialic acid link downward via their 1-position; sialic acid links downward via its 2-position. Split in chromatogram indicates change in extracted ion chromatograms: m/z 117 + 129 for hexose residues and m/z 116 + 158 for N-acetylhexosamine (HexNAc) residues.

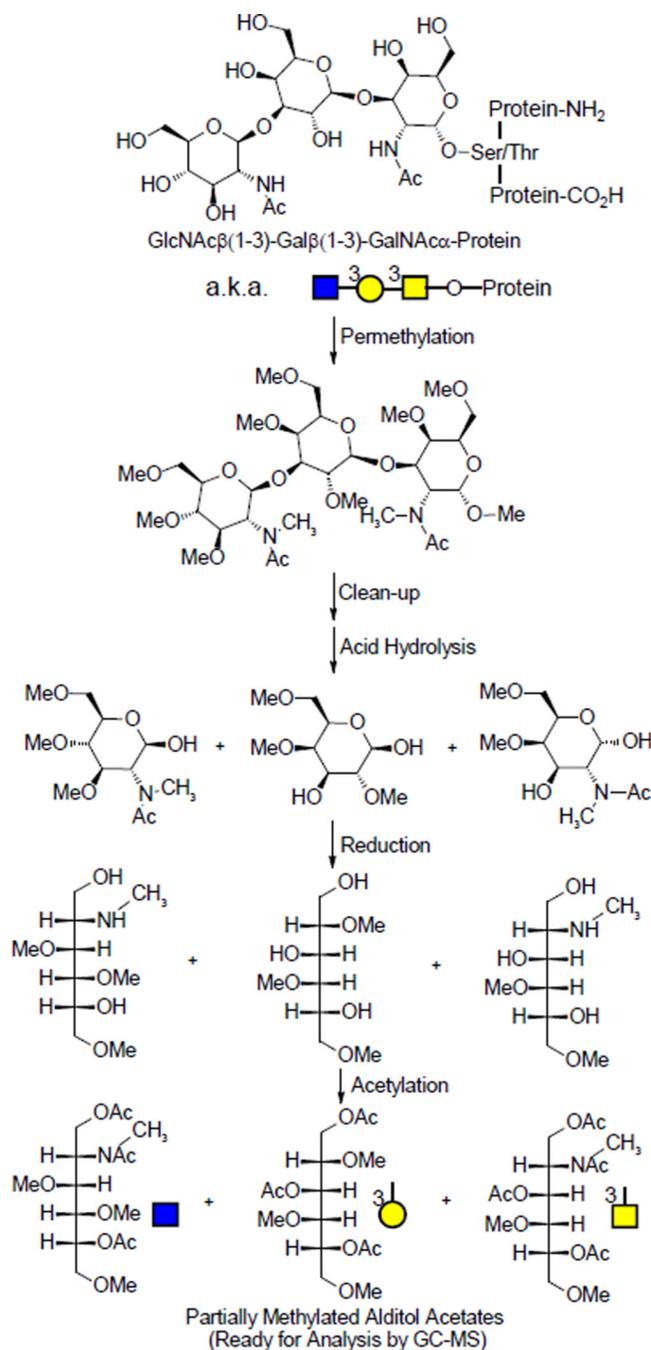


Figure 2. Molecular overview of the global glycan methylation analysis procedure. An O-linked glycan is illustrated; these are released during the permethylation process, which has been adapted from Goetz.²² Following permethylation and hydrolysis, monosaccharides are reduced and nascent hydroxyl groups “marked” by acetylation. The unique pattern of methylation and acetylation in the final partially methylated alditol acetates (PMAAs) corresponds to the unique “glycan node” in the original intact polymer and provides the molecular basis for separation and quantification by GC-MS. N-linked and glycolipid glycans are released as linkage-marked monosaccharides during acid hydrolysis.

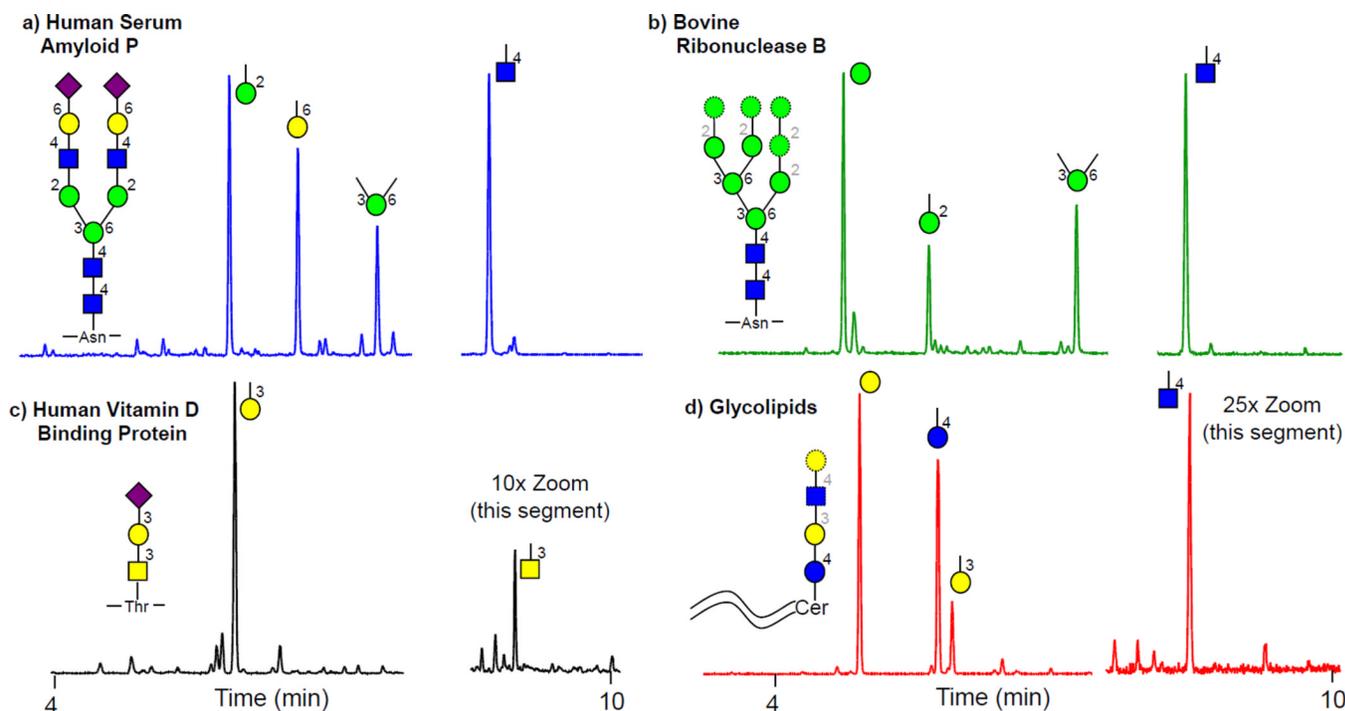


Figure 3.

Evidence that N-linked and glycolipid glycans were captured by the clean-up protocol and subsequently subjected to methylation analysis. This occurred despite the fact that they were not released during permethylation like O-linked glycans²². Top row: **a)** pre-purified Human Serum Amyloid P (contains only one complex-type N-linked glycan⁴⁶⁻⁴⁷), **b)** pre-purified Bovine Ribonuclease B (RNase B, contains only one high mannose N-linked glycan⁴⁸⁻⁴⁹). Bottom row: **c)** pre-purified Human Vitamin D Binding Protein (DBP, contains a NeuNAc2-3Gal1-3GalNAc O-linked glycan and no N-linked glycans⁵⁰⁻⁵¹) **d)** Pre-purified neutral glycosphingolipids from human granulocytes (largely characterized by their lactose (Gal1-4Glc)-base, 3-Gal and 4-GlcNAc-containing structures). Extracted ion chromatograms and symbol legend in this Figure are the same as in Fig. 1. Dotted borders around monosaccharides and greyed out linkage numbers indicate potential heterogeneity in the glycan structure across different protein or lipid molecules.

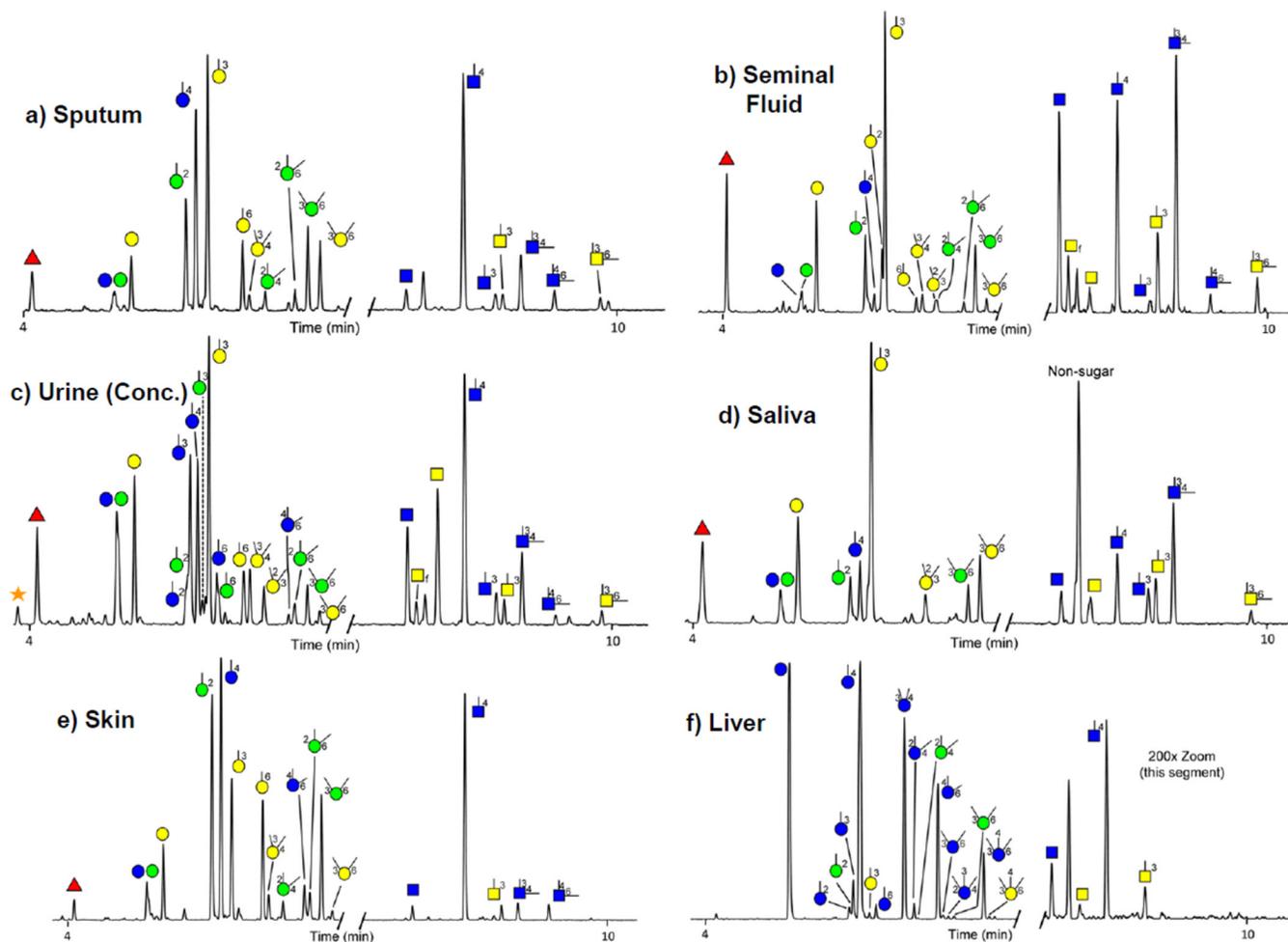


Figure 4.

Illustrative results from biomatrix compatibility studies. The analytical technique may be applied to 10-NL volumes of any whole biofluid or homogenized tissue. Qualitatively diverse results are obtained for **a**) sputum (homogenized), **b**) seminal fluid (without sperm), **c**) urine (concentrated ~10x prior to analysis), **d**) saliva, **e**) skin harvested from an abrasion wound and **f**) liver (bovine). Legend is provided in Fig. 1. The “f” next to terminal t-GalNAc in the urine sample indicates a furanose (5-membered) ring structure, which likely arises from the presence of some structurally interchangeable free GalNAc in the sample. Glycan nodes derived from glycogen dominate the liver sample.

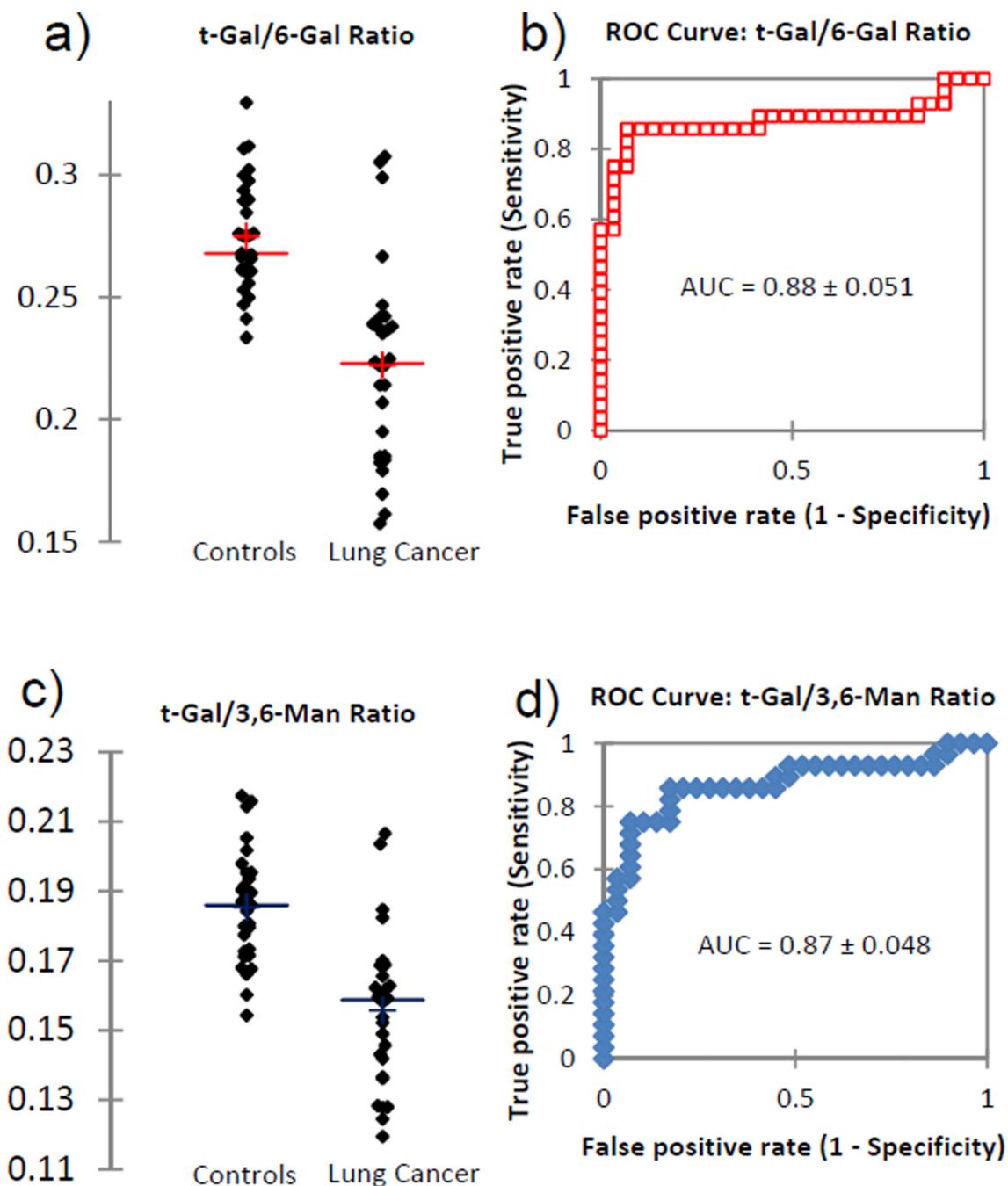


Figure 5. Clinical performance of the top blood plasma-based glycan node ratios (GNRs) in distinguishing newly diagnosed lung cancer patients ($n=28$) from age/gender/smoking status-matched controls ($n=29$). **a)** Univariate distribution of the t-Gal/6-Gal GNR **b)** ROC curve for t-Gal/6-Gal, **c)** univariate distribution of the t-Gal/3,6-Man GNR, **d)** ROC curve for t-Gal/3,6-Gal. For both of these GNRs the same two samples from squamous cell carcinoma patients produced two outliers on the opposite side of the control distribution. Since they were completely separate from the control distribution they were excluded from

the ROC curve analysis. Table 1 summarizes the clinical performance characteristics for the top 12 diagnostic GNRs in lung cancer.

Table 1

Summary clinical information on gender, age and smoking status for the 59 samples analyzed from the Lung Cancer in Central and Eastern Europe study.

	Gender	Age^a	Tobacco Pack Years^b
Controls	15 Male	63.1 ± 7.6	17.3 ± 15.2
	14 Female		
Lung Cancer Cases	16 Male	60 ± 10.7	27.0 ± 20.7
	14 Female		

^aStudent's t-test p-value for Controls vs. Cases = 0.21

^bStudent's t-test p-value for Controls vs. Cases = 0.046

Table 2

Analytical reproducibility and clinical performance characteristics of the 12 topperforming blood plasma-based glycan node ratios in lung cancer. Reproducibility was assessed through the analysis of six samples per batch on three separate days. Diagnostic capacity was maintained in the analysis of smokers only and, with the exception of two node ratios that share a common denominator (3-GalNAc), there was a loss of diagnostic capacity for smokers vs. non-smokers.

Glycan Node Ratio	Intra-assay Precision (%CV)	Inter-assay Precision (%CV)	Cancer (n=28) vs. Non-Cancer (n=29)	Smokers Only: Cancer (n=25) vs. Non-Cancer (n=23)	Trend in Cancer (Increased or Decreased, I/D)	Smokers (n = 48) vs. Non-Smokers (n=9) ^b
			ROC AUC ± SE	ROC AUC ± SE		ROC AUC ± SE
t-Gal/6-Gal	2.34	3.76	0.878 ± 0.051 ^a	0.897 ± 0.048 ^a	D	0.681 ± 0.087 ^c
t-Gal/3,6-Man	4.35	5.69	0.869 ± 0.048 ^a	0.868 ± 0.051 ^a	D	NS ^d
2,4-Man/3-GalNAc	7.76	11.97	0.793 ± 0.055	0.775 ± 0.062	I	0.804 ± 0.057
t-Gal/2,4-Man	7.06	7.38	0.79 ± 0.06	0.789 ± 0.063	D	NS ^d
2,4-Man/3,4,6-Man	8.59	11.13	0.786 ± 0.06	0.775 ± 0.066	I	NS ^d
2-Man/2,4-Man	6.08	8.93	0.78 ± 0.058	0.781 ± 0.062	D	NS ^d
6-Gal/3-GalNAc	5.29	9.56	0.777 ± 0.058	0.778 ± 0.064	I	0.88 ± 0.043
2,4-Man/t-GlcNAc	4.47	5.12	0.772 ± 0.058	0.776 ± 0.062	I	NS ^d
6-Gal/3,4,6-Man	4.61	7.15	0.772 ± 0.06	0.767 ± 0.065	I	NS ^d
2,6-Man/3,4,6-Man	5.24	5.59	0.766 ± 0.063	0.749 ± 0.069	I	NS ^d
3,6-Man/3,4,6-Man	3.35	4.89	0.764 ± 0.061	0.752 ± 0.066	I	NS ^d
t-Gal/2,6-Man	5.9	9.82	0.76 ± 0.064	0.767 ± 0.067	D	NS ^d

^aIncluding two Lung Cancer-group outliers that lie on the *opposite side* of the Control distribution and are *completely separate* from it gives ROC c-statistics (AUCs) of 0.82 ± 0.062 and 0.81 ± 0.060 for the top 2 node ratios, respectively (0.83 ± 0.063 and 0.80 ± 0.064 for the Smokers Only group).

^bRegardless of cancer status.

^cThe difference from the "Cancer vs. Non-Cancer" group is 0.197 ± 0.10 (statistically significant at the 2σ-level).

^dNot statistically significant from an ROC AUC of 0.5