

Smoking and lung cancer-induced changes in N-glycosylation of blood serum proteins

Jacqueline A Vasseur^{2,3}, John A Goetz^{2,3}, William R Alley Jr^{2,3}, and Milos V Novotny^{1,2,3,4}

²Department of Chemistry; ³National Center for Glycomics and Glycoproteomics, Indiana University, Bloomington, IN 47405, USA; and ⁴Indiana University School of Medicine, Indianapolis, IN 46202, USA

Received on February 6, 2012; revised on June 18, 2012; accepted on July 4, 2012

Glycosylation is a key post-translational protein modification which appears important in malignant transformation and tumor metastasis. Abnormal glycosylation of different proteins can often be measured in the blood serum. In this study, we extend our serum-based structural investigations to samples provided by patients diagnosed with lung cancer, paying particular attention to the effects of smoking on the serum glycomic traces. Following a battery of glycomic tests, we find that several fucosylated tetra-antennary structures with varying degrees of sialylation are increased in their abundances in control samples provided by the former smokers, with further elevations in the lung cancer patients who were former smokers. Further detailed investigations demonstrated that the level of outer-arm fucosylation was elevated in the control samples of the former smokers and again in the lung cancer samples provided by the former smokers. This trend was particularly noticeable for the tri- and tetra-antennary structures. Different ratios of sialylation linkages were also observed that could be correlated with the different states of health and smoking status. Decreases in the abundance levels of isomers with two and three α 2,3-linked sialic acids and an increased abundance of an isomer with two α 2,6-linked sialic acids were noted for a fucosylated tri-sialylated tri-antennary glycan. These results demonstrate the long-term effects of smoking on glycomic profiles and that this factor needs to be considered in these and other serum-based analyses.

Keywords: disease markers / glycomics / non-small cell lung carcinoma / permethylation / smoking

Introduction

Human blood serum is an abundant source of glycoproteins with different structures, origins and biological functions. Since it has long been appreciated that different tissue types, including cancer tissues, may feature differential expressions of glycoprotein carbohydrate structures in health and disease, some cancer-related glycoproteins may become minor components of the circulatory system through cell rupture events or the shedding of cell-surface glycoproteins. In a search for cancer diagnostic biomarkers, such indicative glycoprotein analytes may seem to be attractive primary targets, even though the tumor-associated proteins are expected to be present at only pg/mL concentration ranges (Anderson and Anderson 2002), making their direct glycomic measurements currently difficult to perform.

While many other serum glycoproteins could be depleted (Lin et al. 2009; Andersen et al. 2010; Smith et al. 2011; Zeng et al. 2011) from a sample in order to enhance the measurements of “true biomarkers”, glycosylation changes in the more abundant glycoproteins, such as in certain proteins with immune functions and the so-called acute-phase proteins (Salдова et al. 2007, 2008; Arnold et al. 2008; Gornik and Lauc 2008), may also be informative of different disease conditions by themselves. We have previously shown the approach of serum glycomic profiling through biomolecular mass spectrometry (MS) to be useful in distinguishing control individuals from patients diagnosed with breast (Kyselova et al. 2008), ovarian (Alley, Vasseur, et al. 2012), liver (Tang et al. 2010; Kang et al. 2011), esophageal (Mechref et al. 2009) and metastatic prostate (Kyselova et al. 2007) cancers. Along with determining the potential utility of glycomic procedures to monitor disease progression/regression and the effectiveness of different treatments, the goal of such methods is to evaluate the structural differences between controls and the patients diagnosed with different types of cancer. Based on the analyses of N-linked glycans, we have been able to achieve a certain level of specificity in the measurements of the location of fucose residues as being on either a glycan’s outer arm or the core in different cancers (Alley, Vasseur, et al. 2012; Alley in preparation).

To investigate further the common glycosylation features and variations in the different cancer conditions, we have extended our analyses to lung cancer in the study described in this communication. Across the globe, lung cancer is the most common malignancy that leads in cancer-related deaths (Boyle and Levin 2008). Each year in the USA alone, 213,000 new

¹To whom correspondence should be addressed: Tel: +812-855-4532; Fax: +812-855-8300; e-mail: novotny@indiana.edu

cases of lung cancer are diagnosed; 80% of those are classified as non-small cell lung carcinoma (NSCLC), and 160,000 deaths annually are attributed to this disease (Mulshine and Sullivan 2005). Similar to many other malignancies, lung cancer is frequently diagnosed in its later stages and with a bleak prognosis: the 5-year survival rate for individuals diagnosed with late-stage lung cancer is less than 20% (Okada et al. 2005; Kassis et al. 2009). However, when detected early, with the surgical removal of the tumor remaining an option, the 5-year survival rate improves to 86% (Okada et al. 2005; Kassis et al. 2009). The early detection of lung cancer is a key to improving a patient's overall prognosis and has been the focus of a significant amount of research.

Although no general screening procedure for lung cancer has been recommended or is being used clinically (Okada et al. 2005; Ostroff et al. 2010), progress has been made for its detection through the use of imaging techniques. A recent survey of individuals with a high risk for developing lung cancer by the National Institutes of Health demonstrated that the patients screened through low-dose helical computed tomography (CT) had a 20% lower mortality rate than those screened by a traditional chest x-ray, presumably due to the improved image quality associated with the CT (Aberle et al. 2011). However, the relatively high cost of the helical CT, a limited access to this technology in some geographical areas, and problems of differentiating benign pulmonary nodules from lung cancer (Welch et al. 2007; Wilson et al. 2008), leading to higher-than-desired false-positive rates, have limited its widespread use.

Molecular tests based on glycoconjugate measurements have also been developed to identify potential markers of lung cancer (Heo et al. 2007; Ueda et al. 2009; Zeng et al. 2010; Arnold et al. 2011). In one study, glycopeptides were enriched using hydrazide beads, which unfortunately required a disassociation of the glycans from their peptides prior to their mass-spectral detection and consequently the detailed information pertaining to the potential site-specific changes to carbohydrate structures could not be obtained. However, the same study demonstrated that 38 glycopeptides originating from 22 unique glycoproteins were differentially expressed in their abundance levels in the lung cancer serum samples when compared with controls (Zeng et al. 2010). An additional study utilizing multiple-lectin affinity chromatography suggested that 38 serum proteins were altered in their abundance in the pathological samples (Heo et al. 2007). Interestingly, plasma kallikrein protein fragment was detected in 25 of the 28 disease samples, but was observed in only one of the eight control samples (Heo et al. 2007). Other methods of analyzing glycoconjugates in lung cancer studies have included a coupling of the lectins of jacalin and *Sambucus nigra* agglutinin to Proteinchips, which have been used to further reveal reductions in the levels of α 2,6-linked sialic acid linkages in the glycans associated with apolipoprotein C-III (Ueda et al. 2009). Although these approaches have examined glycosylation patterns on proteins, released carbohydrates have been analyzed in a recent glycomic study, which has further implicated several structures as being altered in their relative abundances, including increases in the mono-antennary structures and tri-sialylated glycans in the serum of lung cancer patients.

Additionally, increases in sialyl Lewis x epitopes were observed (Arnold et al. 2011). Many of these changes appeared to be correlated with the stage of the disease.

In this study, we have subjected microliter volumes of serum provided by control individuals ($n=30$) and patients diagnosed with NSCLC ($n=46$) to a battery of glycomic tests, including our standard N-linked profiling of permethylated glycans, an analysis of the location of fucose residues, and a determination of the sialic acid linkages and their possible alterations due to the pathological condition. As smoking has been indicated to be one of the most significant factors leading to lung cancer and has been shown to significantly alter glycomic profiles through different analytical methods (Knezevic et al. 2010; Arnold et al. 2011), we have further categorized the samples according to the smoking history of the individuals enrolled in this study. The mass-spectral data were then statistically evaluated by a one-way analysis of variance (ANOVA) to determine the significance of the differences between the various sample sets. The data for those oligosaccharides, which passed this test (with a P -value <0.05), were further subjected to a receiver operator characteristics (ROC) test to determine their diagnostic potential according to their area under the curve (AUC) values.

Results

In this study, 46 serum samples that were provided by the patients diagnosed with NSCLC and 30 samples given by individuals, which were used as control samples, were subjected to a series of glycomic tests to determine possible alterations to the abundance levels of the various glycans. These tests included a comprehensive N-linked profiling of permethylated glycans, in addition to an analysis of the location of the fucose residues based on an exoglycosidase digestion performed with a non-specific sialidase and a β -galactosidase. The influence of the different states of health on the linkage ratios of sialic acids was also investigated. As a part of the sample banking protocol, the smoking history of each subject was recorded and the control sample cohort and the pathological sample set were further subdivided according to the smoking histories of the subjects for this study. In the control sample set, 11 subjects were known to be former smokers and 17 individuals had no smoking history, whereas 29 of the lung cancer patients were former smokers. After performing the statistical analyses on the control and cancer groups without considering the subjects' smoking history, the statistical calculations were repeated on these sample groups based on the individual's smoking status to examine its influence on the glycomic patterns.

Typical N-linked glycomic profiles

The blood serum samples yielded glycomic profiles that consisted of ~ 55 unique m/z signals corresponding to the known glycan structures. Traditionally, these individual glycans fall into the broad categories including high-mannose-type glycans, hybrid oligosaccharides and complex structures, which may be further subcategorized as based on their different structural features (i.e. the presence of a "bisecting" *N*-acetylglucosamine (GlcNAc), the extent of glycan

branching or their levels of fucosylation and/or sialylation). Representative N-linked glycomic profiles acquired for a non-smoking control individual, a trace for a control individual who formerly smoked and that of a former smoker diagnosed with NSCLC are shown in Figure 1A–C, respectively. The matrix-assisted laser desorption/ionization (MALDI) tandem MS spectra confirming the structures associated with many of these signals are given as Supplementary data, Figure S1. Among the most prominent changes are those occurring for the *N*-glycans producing the more intense ion signals. For example, the mono-galactosylated bi-antennary core-fucosylated structure, observed at an *m/z* value of 2056.1, was present with the highest intensity in the control sample with no smoking history and appeared to be successively lowered in its ion signal strength in the control individual who was a former smoker, and finally, in the NSCLC patient with a past smoking history. The tri-antennary tri-sialylated structures, both a-fucosylated and fucosylated (observed at *m/z* values of 3618.8 and 3792.9, respectively) also appeared to follow a trend. These oligosaccharides appeared to produce the lowest ion intensities in the non-smoking control sample and were progressively increased in their intensities in the control sample provided by a former smoker, and the patient who was a former smoker diagnosed with NSCLC. Additionally, several lower-intensity signals also seemed to be altered in the pathological samples, as shown in the insets of Figure 1, which depicts the high-mass region of the spectrum where the tri- and tetra-antennary structures, with varying degrees of sialylation and fucosylation, reside. Interestingly, the signals associated with these structures appeared to be increased steadily from the control samples of the individual who never smoked, to control sample of the former smoker and to the sample provided by NSCLC patient with a past history of smoking.

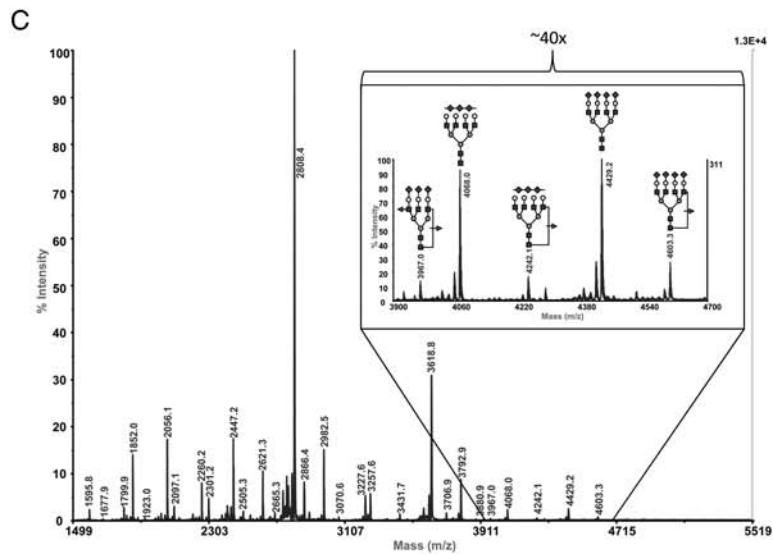
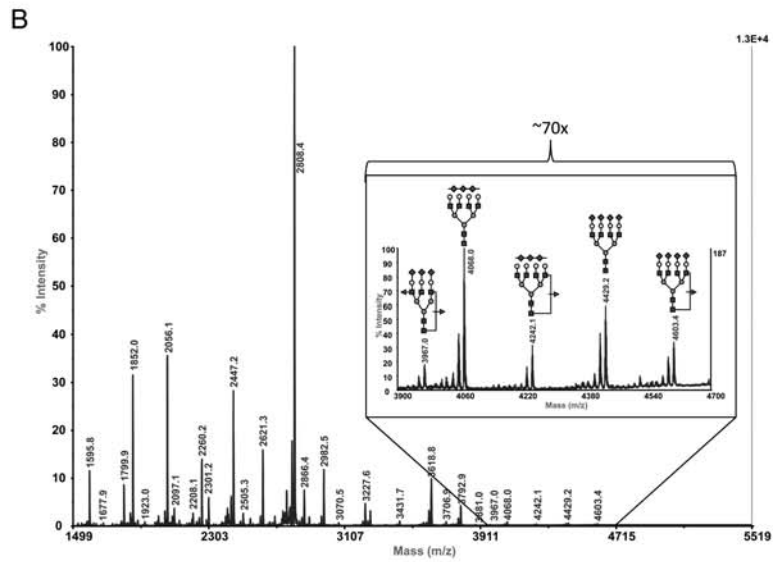
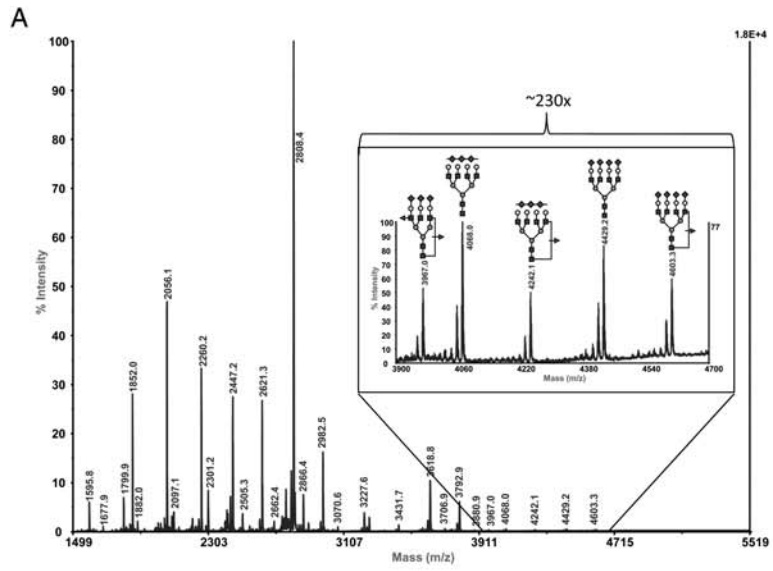
Analysis of glycan changes due to NSCLC

As the first step in observing the possible differences between the control and pathological cohorts, without considering the possible effects of smoking on the glycomic profiles, a principal component analysis (PCA) was performed. This type of analysis reduces the data sets with higher dimensions into a series of orthogonal principal components that retain much of the original information, with the first two principal components (PC1 and PC2, respectively) generally retaining the most data. A plot of PC1 vs PC2 is shown in Figure 2. Although there is a small degree of overlap within the two sample sets, particularly around the origin of the plot, the majority of the control samples are found in the quadrants to the left of the origin, while the pathological samples are mainly located to the right. A fairly high degree of separation between the different sample sets in PC1 indicates that there are likely significant differences between their carbohydrate patterns.

To reveal which types of glycans contributed to the differentiation of the sample cohorts observed in the PCAs, the glycan pool was subdivided into 14 different classes, as based on their distinct structural features, e.g. high-mannose-type structures, sialylated-only carbohydrates or fucosylated-and-sialylated oligosaccharides. The statistical evaluations were then performed on the sum of the normalized intensities

of the constituents for each group. The results of these calculations are listed in Supplementary data, Table SI, which shows that 8 of the 14 groups resulted in acceptable statistical values. (In this part of the study, only the normalized glycomic data that resulted in a *P*-value <0.05 and an AUC score higher than 0.7 or lower than 0.3 were considered to be good candidates for a further study.) The high-mannose class (average relative intensities of 5.9 and 4.8% in the control and pathological sample sets, respectively), the neutral fucosylated class (average relative intensities of 24.4 and 17.9% in the control and lung cancer cohorts, respectively), and the neutral galactosylated class (average relative intensities of 16.9 and 10.3% in the control and disease samples, respectively) appeared to be lowered in their relative abundances in the pathological samples, all with *P*-values substantially <0.05 and with AUC scores in the mid-0.7 range for the high-mannose and the neutral fucosylated classes and in the high 0.8 range for the neutral galactosylated class. The total “sialylated-only” class (average relative intensities of 51.2 and 57.4% in the control and pathological sample sets, respectively), including the total bi-antennary class (45.2 and 50.0% for the control and NSCLC samples, respectively) and the total fucosylated-and-sialylated class (average relative intensities of 9.04 and 11.1% in the control and pathological sample sets, respectively), including its tri- (average relative intensity of 1.91% in the control samples and 3.81% in the cancer samples) and tetra-antennary (0.159 and 0.390% relative intensities in the control and pathological samples, respectively) subclasses, were elevated in their abundance levels in the lung cancer sample set. The data associated with these particular groups of analytes also resulted in *P*-values ranging from 10^{-4} to 10^{-8} , whereas their AUC scores were in the mid-0.7 range for the total sialylated-only group and the fucosylated-and-sialylated class. The information associated with tri- and tetra-antennary subclasses of the fucosylated-and-sialylated group both generated AUC scores in the high 0.8 range. The altered abundance levels for the high-mannose class, the total neutral fucosylated group, the total neutral galactosylated class of oligosaccharides, the total sialylated-only glycans, the total fucosylated-and-sialylated group and its tri- and tetra-antennary subclasses are shown as the notched-box plots in Figure 3A–G, respectively.

The statistical tests were then performed on the data acquired for the specific *m/z* values to determine which particular glycan structures were altered in their abundance levels and thus may be used to distinguish the different sample sets. These evaluations showed that data generated by 29 of the 55 *m/z* values, corresponding to the known glycan structures, resulted in significant statistical information, the results of which are presented in Table I. Among the glycans, most significantly altered in their expression levels were several smaller structures observed at *m/z* values of 2056.1, a core-fucosylated mono-galactosylated bi-antennary structure, with an *m/z* value of 2086.1, a di-galactosylated bi-antennary glycan, and an *m/z* value of 2260.2, which corresponds to a core-fucosylated di-galactosylated bi-antennary oligosaccharide. All of the average intensities for all of these structures were significantly suppressed by nearly 50% in the pathological samples. The structures giving rise to *m/z* values of



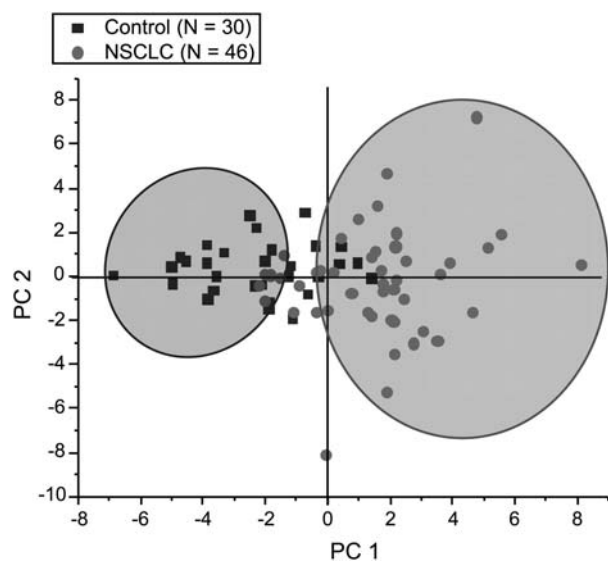


Fig. 2. Score plot of the first two principal components of a PCA of control samples for NSCLC samples without a consideration of the smoking status of the individuals.

2056.1 and 2260.2 both produced fairly intense ions, with average normalized intensities of 8.9 and 3.6% in the control samples, respectively, and only 5.5% and 1.7% in the pathological samples, respectively. The structure producing the ion observed at the m/z value of 2086.1 generally yielded less intense ion signals, 0.710% of the total intensity that was associated with this ion in the control samples and only 0.442% in the lung cancer samples. Interestingly, our prior work has implicated all of these structures as being associated primarily with IgG (Alley, Vasseur, et al. 2012; Svoboda et al. 2012). They have been observed as suppressed in their expression levels in different cancers that we had studied previously (Kyselova et al. 2007, 2008; Alley, Vasseur, et al. 2012).

Several of the key glycans that were observed to be increased in their expression levels routinely in our past serum-based cancer investigations (Kyselova et al. 2007, 2008; Kang et al. 2011; Alley, Vasseur, et al. 2012) are the tri- and tetra-branched structures with varying levels of sialylation and fucosylation. The same trend was also observed in the lung cancer samples analyzed in this study. While the tri-sialylated tri-antennary glycan, observed at an m/z value of 3618.8, failed to produce acceptable statistical figures, its fucosylated analog, producing the ion recorded at an m/z value of 3792.9, was significantly elevated in the lung cancer samples, indicating the importance of fucosylation in cancer. This fucosylated structure was observed with an average relative intensity of 1.07% in the control samples and was elevated nearly three times in the lung cancer samples to 3.18%. The statistical calculations resulted in a P -value of 3.21×10^{-8} and an accurate AUC score of 0.858. Additionally, a doubly

fucosylated tri-sialylated tri-antennary glycan, appearing at an m/z value of 3967.0, was often absent in the control spectra, resulting in an average relative intensity of only 0.054%. However, it was observed in the majority of the cancer samples, contributing 0.124% on average to the total intensity in the lung cancer cohort. The statistical comparison of the data for this structure produced a P -value of 1.10×10^{-6} and an AUC score of 0.849.

Perhaps the best example of a tetra-antennary structure with a significantly altered expression level is that observed at the m/z value of 4603.3, corresponding to a fucosylated tetra-sialylated tetra-antennary structure. In the control samples, this particular glycan contributed only 0.051% to the total intensity, while in the pathological samples, its relative abundance was increased to 0.149%. The statistical comparisons of the data for this structure produced a P -value of 4.45×10^{-7} and an AUC score of 0.895. Very similar observations were made for the tri-sialylated version of this structure (Table I).

Effects of smoking history on glycan profiles

Recent results from a different laboratory using an alternative methodology have shown that smoking can induce an apparent acute-phase response (Knezevic et al. 2010; Arnold et al. 2011). The differences appeared to be particularly noticeable for a number of tetra-antennary structures. As a part of the clinical documentation for the samples used in study presented in this communication, the subjects' smoking histories were recorded. Using this information, we have subdivided our samples into the sets composed of the control samples provided by individuals who had no prior history of smoking, control samples donated by the individuals who formerly smoked for an average of 9.7 years and lung cancer patients who were former smokers to examine the possible effects of smoking as revealed by our analytical methods. Unfortunately, we had insufficient numbers for meaningful statistical analyses for both control patients who currently smoke and the NSCLC patients who had no prior history of smoking.

Using these new sample groups based on the smoking histories of the subjects, we once again performed a PCA; the score plot for the first two components of this test is shown in Figure 4. In this plot, the control samples provided by the individuals with no smoking history tended to cluster mostly in the lower left quadrant of the plot, whereas the NSCLC patients who were former smokers were primarily located to in the quadrants to the right of the origin. In the immediate vicinity of the origin, some intermixing of these two states of health has apparently occurred. However, it was in this region that the majority of the control samples collected from the former smokers clustered together. These particular samples were fairly well separated in the PC2 dimension from the non-smoking control samples. This seems to indicate that our methods are also capable of distinguishing smokers from non-smokers and that these types of samples appeared as

Fig. 1. Representative MALDI MS-based glycomic profiles for (A) a control patient who never smoked; (B) a control sample provided by a former smoker and (C) an NSCLC patient who was a former smoker. The insets depict the higher mass regions of the spectra, where the tetra-antennary oligosaccharides, present as fucosylated and a-fucosylated structures, are found. These analytes appeared to produce the most intense signals in the lung cancer sample, and the lowest intensities in the control sample provided by an individual with no smoking history. These structures in the control sample given by a former smoker appeared to possess intermediate values. Symbols: square, GlcNAc; circle, mannose; yellow, galactose; diamond, N-acetylneuraminic acid; triangle, fucose.

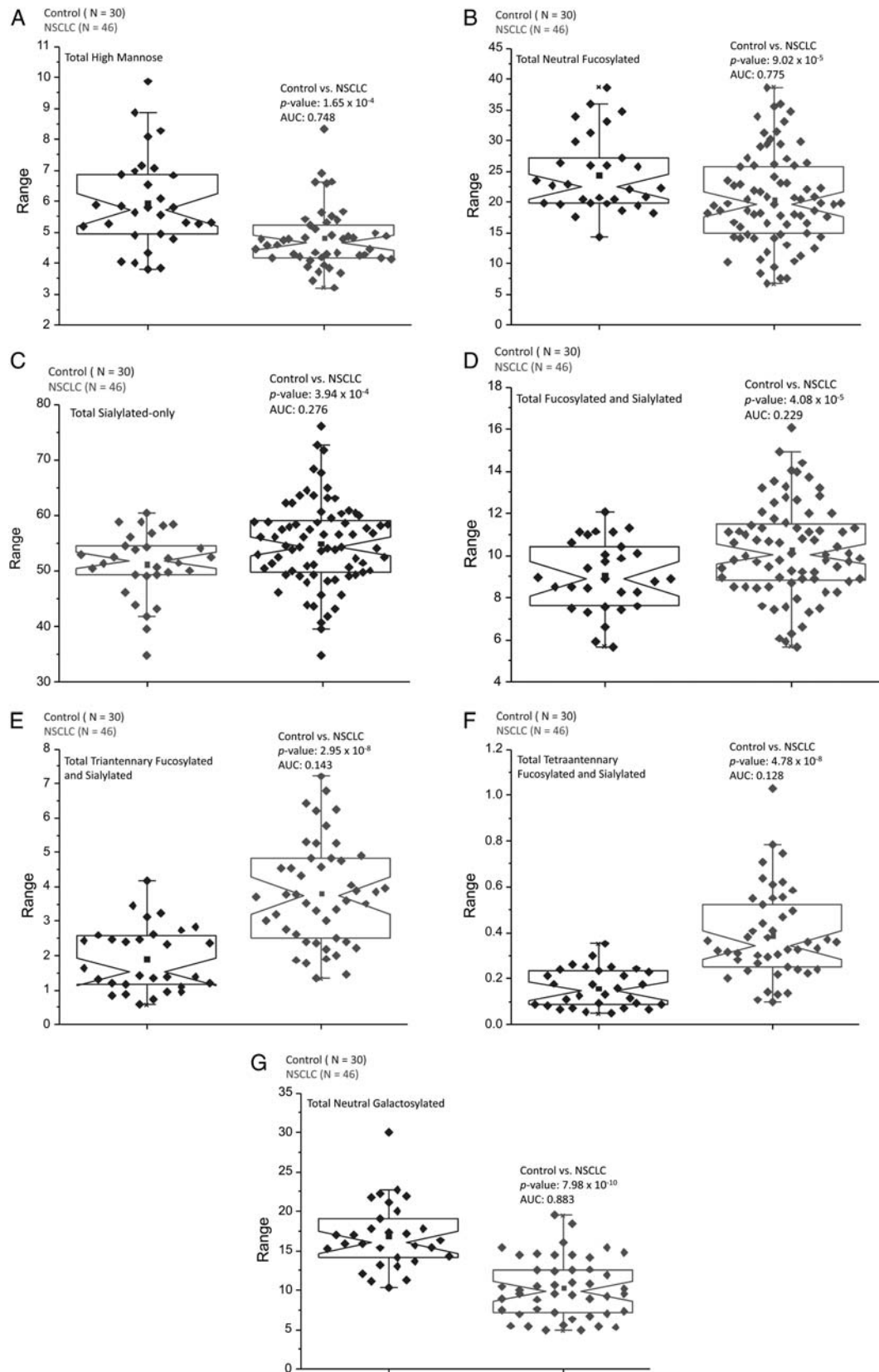


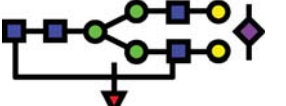

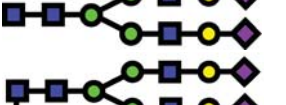
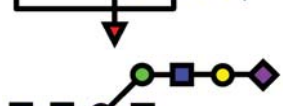
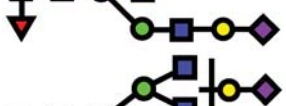
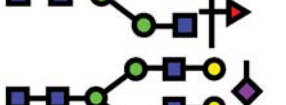
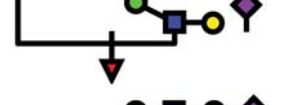
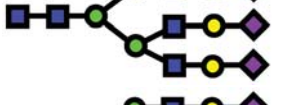
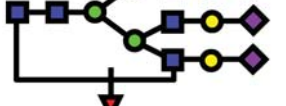
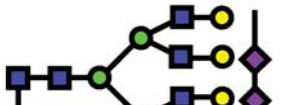

Fig. 3. Changes in the abundance levels of particular classes of glycans between control samples and NSCLC samples. Significant alterations were observed for (A) high-mannose class of glycans; (B) neutral fucosylated family; (C) total sialylated group; (D) fucosylated and sialylated group; (E) tri-antennary fucosylated and sialylated glycans; (F) tetra-antennary fucosylated and sialylated group of oligosaccharides and (G) total set of neutral galactosylated structures.

Table I. Average relative intensities, standard error of the mean (SEM) values, *P*-values and AUC scores for glycans with a *P*-value of <0.05

Structure	<i>m/z</i>	Control		Lung cancer		<i>P</i> -value	AUC
		Avg Rel Int	SEM	Avg Rel Int	SEM		
	1595.9	1.75	0.123	1.29	0.061	5.18×10^{-4}	0.264
	1636.8	0.101	0.005	0.076	0.003	1.25×10^{-5}	0.225
	1799.9	1.86	0.098	1.39	0.049	1.61×10^{-5}	0.227
	1882.0	0.634	0.162	0.289	0.025	0.012	0.212
	2004.0	0.447	0.022	0.337	0.016	1.21×10^{-4}	0.217
	2015.1	0.098	0.007	0.082	0.005	0.038	0.361
	2026.0	0.100	0.005	0.067	0.004	7.11×10^{-7}	0.173
	2056.1	8.90	0.44	5.50	0.403	5.90×10^{-8}	0.151
	2086.1	0.710	0.056	0.442	0.019	1.25×10^{-6}	0.162
	2260.2	3.58	0.283	1.72	0.149	1.63×10^{-8}	0.107
	2301.2	1.83	0.068	1.37	0.115	1.29×10^{-3}	0.217
	2331.2	0.138	0.006	0.110	0.005	3.43×10^{-4}	0.252
	2417.2	0.599	0.022	0.472	0.015	1.21×10^{-6}	0.188
	2488.3	0.277	0.025	0.366	0.027	1.600×10^{-2}	0.658
	2505.3	0.696	0.031	0.518	0.024	1.51×10^{-5}	0.199

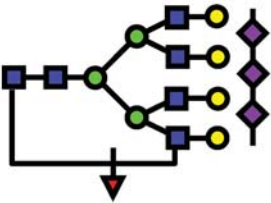
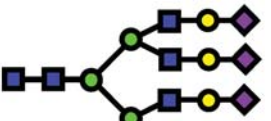
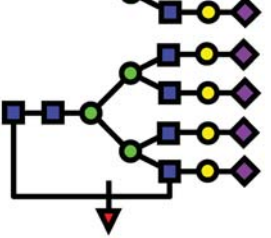
Continued

Table I. (Continued)

Structure	<i>m/z</i>	Control		Lung cancer		<i>P</i> -value	AUC
		Avg Rel Int	SEM	Avg Rel Int	SEM		
	2621.3	3.44	0.156	2.75	0.13	1.24×10^{-3}	0.254
	2662.4	0.188	0.010	0.028	0.034	2.550E-02	0.675
	2808.4	36.9	0.877	41.9	0.811	1.16×10^{-4}	0.754
	2982.5	3.53	0.142	4.13	0.157	4.77×10^{-3}	0.679
	3227.6	1.01	0.081	1.33	0.091	0.015	0.646
	3401.7	0.024	0.002	0.040	0.002	2.88×10^{-6}	0.838
	3431.7	0.257	0.018	0.427	0.023	2.06×10^{-7}	0.823
	3618.8	3.93	0.24	4.89	0.328	0.038	0.624
	3792.9	1.47	0.148	3.19	0.201	3.21×10^{-8}	0.858
	3881.0	0.049	0.004	0.099	0.007	2.32×10^{-7}	0.838
	3967.0	0.054	0.007	0.124	0.01	1.10×10^{-6}	0.849

Continued

Table I. (Continued)

Structure	<i>m/z</i>	Control		Lung cancer		<i>P</i> -value	AUC
		Avg Rel Int	SEM	Avg Rel Int	SEM		
	4242.1	0.059	0.006	0.143	0.011	8.89×10^{-8}	0.877
	4429.2	0.119	0.013	0.190	0.021	6.33×10^{-3}	0.686
	4603.3	0.051	0.005	0.149	0.014	4.45×10^{-7}	0.895

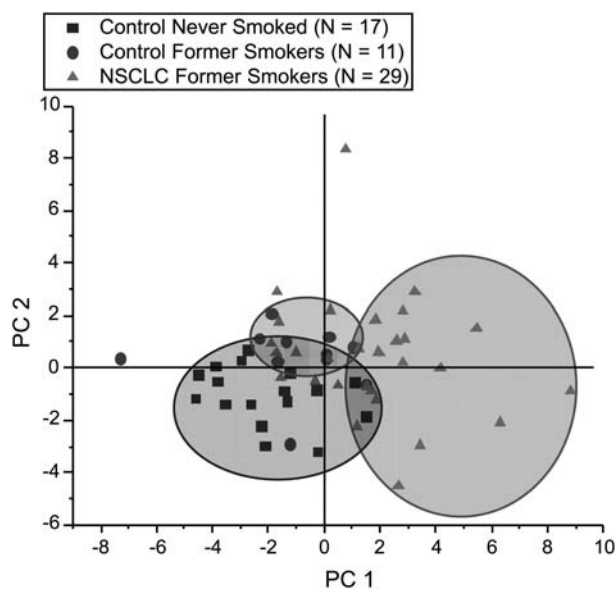


Fig. 4. Plot of the first two principal components for a PCA of the control samples provided by individuals who never smoked, control samples donated by former smokers and for NSCLC samples given by former smokers.

intermediates in our data. Therefore, it seems that the overall study design needs to account for the smoking history of its subjects.

As in the previous analyses without a consideration for the smoking history of the individuals, we now grouped the glycans into the same 14 classes based on their structural features and performed the usual statistical tests, the results of which are summarized in Supplementary data, Table SII.

After performing these calculations, the most notably altered classes between the control samples provided by the individuals with no smoking history and those who were former smokers were fucosylated-and-sialylated family, with the tri- and tetra-antennary subgroups being the most significantly affected, as shown in Figure 5A–C. This figure presents the notched-box plots for the total fucosylated-and-sialylated class of structures and its tri- and tetra-antennary subclasses, respectively. The total fucosylated-and-sialylated class of glycans contributed $\sim 8.3\%$ to the total intensity in the control samples given by patients with no smoking history and just over 10% in the sample set comprised of formerly smoking control samples. These data resulted in a *P*-value of 5.23×10^{-3} and an AUC score of 0.829. However, when the control samples provided by former smokers were statistically compared with the lung cancer samples given by former smokers, these types of structures did not seem to be significantly different, as based on the *P*-value of 0.147 and an AUC score of 0.627.

Elevated carbohydrate levels which may be associated with smoking were observed for the tri-antennary subclass of the total fucosylated-and-sialylated family, whose abundances were nearly doubled in its value in the control samples obtained from former smokers. These structures contributed 1.43% to the total intensity in the control samples provided by the individuals with no smoking history and 2.7% in those who were former smokers, while the resulting statistical evaluations of these data returned a *P*-value of 1.99×10^{-4} and an AUC score of 0.888. In this subclass, it appears that these types of structures may be slightly elevated with the lung cancer patients who were former smokers when compared with the former smokers in the control group. In these types of lung cancer samples, the tri-antennary structures accounted

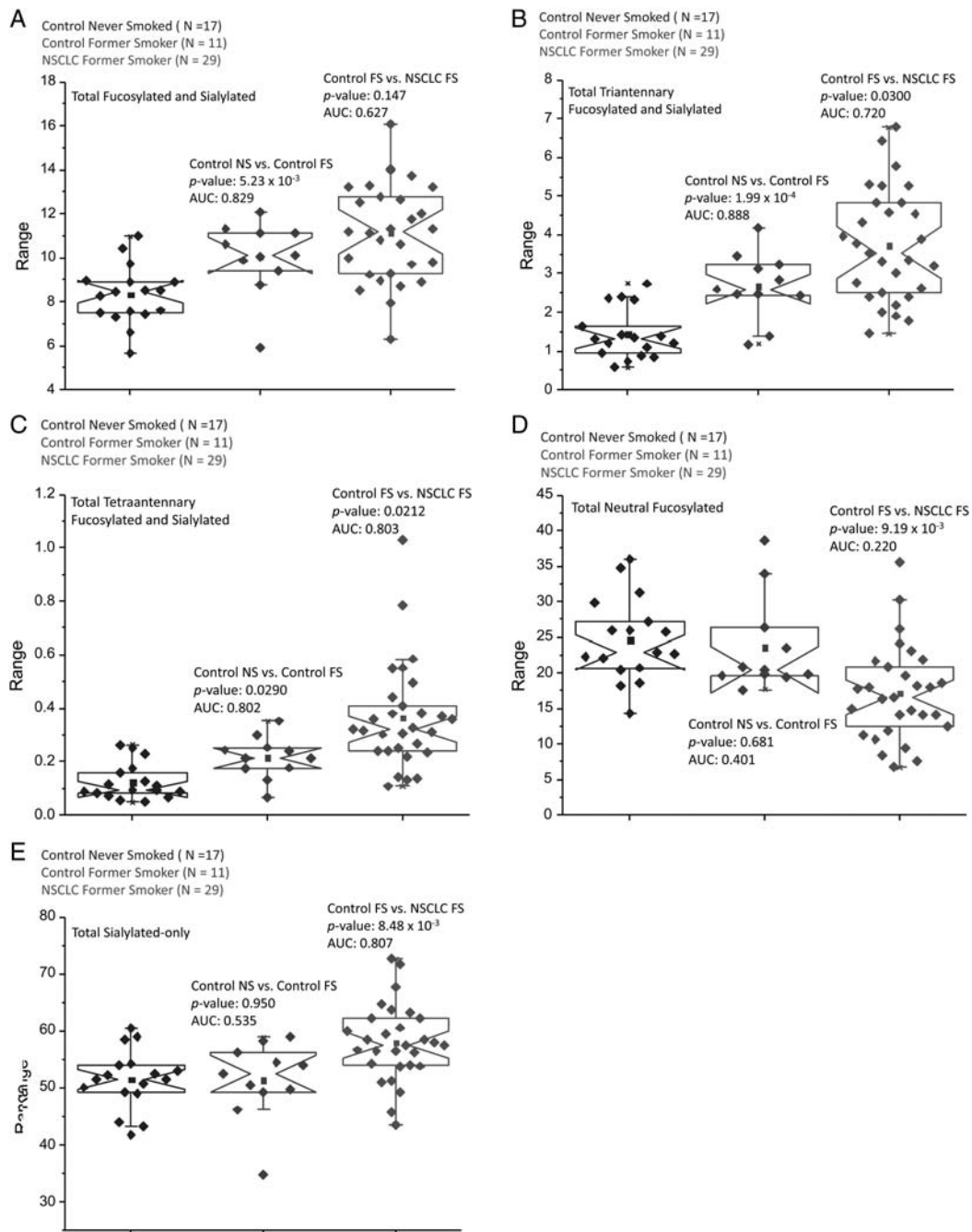


Fig. 5. Notched-box plots depicting the classes of glycans that were significantly different in their abundances based on the smoking histories of the subjects enrolled in this study. When the control samples provided by individuals who never smoked were compared with those of former smokers, statistically supported alterations were observed for (A) the total fucosylated-and-sialylated class of glycans; (B) the tri-antennary subclass and (C) the tetra-antennary subclass of the fucosylated-and-sialylated class. Additionally, (D) the total neutral fucosylated and (E) total sialylated-only classes appeared to be different in their abundances when the control samples of former smokers were compared to the samples given by former smokers diagnosed with lung carcinoma.

for 3.7% of the total intensity and the statistical assessments of the groups composed of former smokers resulted in a P -value of 0.0300 and an AUC value of 0.720. A similar trend was also noted for the tetra-antennary subclass. Though generally observed at much lower intensities than their tri-antennary counterparts, these types of analytes were also seen to be approximately twice as abundant in the control samples

donated by the former smokers, contributing 0.215% to the total intensity, while in the control samples provided by individuals with no smoking history, these glycans accounted for only 0.123% of the total intensity. A statistical evaluation of the data for the two different types of control samples resulted in a P -value of 0.029, whereas the AUC score was determined to be 0.802. Interestingly, this subclass of structures was further

elevated in the lung cancer patients who were former smokers, providing 0.363% to the total intensity and a comparison between the different states of health for the former smokers resulted in a P -value of 0.0212 and an AUC value of 0.803.

Additionally, the total neutral fucosylated class of glycans appeared to present at about the same level in both control sample sets (24.6 and 23.6% for the individuals who never smoked and those who were former smokers, respectively), resulting in a P -value of 0.681. However, in the lung cancer cohort of former smokers, these structures accounted for only 17.1% of the total intensity. A comparison of the different former-smoker sample sets produced a P -value of 9.19×10^{-3} and an AUC score of 0.220. Additionally, the total “sialylated-only” class also appeared to be capable of differentiating the different former-smoker’s states of health, but not the different control sets. In both control pools, these structures were present with nearly identical relative intensities, $\sim 51.5\%$. However, in the lung cancer patients who were former smokers, this class of oligosaccharides accounted for 57.9% of the total intensity. When the former-smoking control samples were compared with these lung cancer samples, the P -value was 8.48×10^{-3} and the AUC score was 0.807. The notched-box plots for the neutral fucosylated class and the sialylated-only structures are shown in Figure 5D and E, respectively.

The specific structures that were capable of distinguishing the two control groups based on the smoking history of its members were primarily fucosylated tri- and tetra-antennary structures with varying degrees of sialylation. Both of the fucosylated tri-antennary structures, the di-sialylated oligosaccharide, represented by an m/z value of 3431.7, and the tri-sialylated glycan, recorded at an m/z value of 3792.9, were observed to be increased in their relative abundances, with the tri-sialylated carbohydrate being nearly doubled in its relative abundance in the control samples provided by the former smokers. The data generated for this carbohydrate resulted in very good statistical figures; a P -value of 1.19×10^{-4} was calculated along with an AUC score of 0.903. This particular structure also appeared to be further increased in the lung cancer patients who were former smokers, though the statistical verification was not as conclusive with a P -value of 0.031 and an AUC score of 0.705. The fucosylated di-, tri- and tetra-sialylated tetra-branched structures, observed at m/z values of 3881.0, 4242.1 and 4603.3, respectively, were also capable of distinguishing the different control sample sets. These structures were also nearly doubled in their abundance values in the control samples donated by the former smokers and continued to be increased in the cancer samples provided by the former smokers. The statistical comparisons between the control groups for these structures resulted in acceptable P -values in the 10^{-3} range and AUC scores ranging from the high 0.7’s to low 0.8’s. When the data collected for these glycans were compared for the control set comprised of the samples acquired from the smokers to those given by the lung cancer patients who had stopped smoking, the P -values ranged from 0.027 to 0.031, with AUC scores being between 0.72 and 0.800. A potentially key structure may be the tetra-sialylated tetra-branched glycan, appearing at an m/z value of 4429.2. This analyte seemed to be present with different statistically supported abundance levels in each of the three

different sample sets. When the two different control sample sets were compared, a P -value of 1.78×10^{-3} was calculated, along with an AUC score of 0.828. When the control samples of former smokers were compared with the cancer patients with a past smoking history, the P -value was calculated to be 0.027 and the AUC score was 0.759.

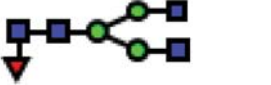
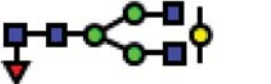
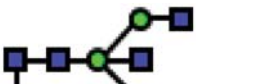

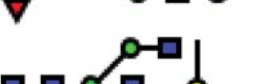

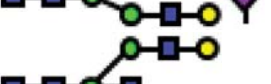
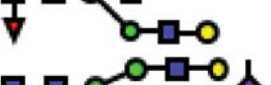
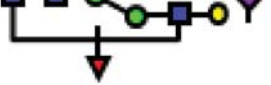
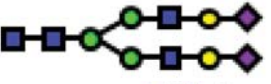
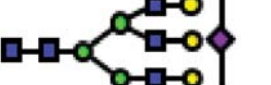
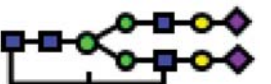

Additionally, the neutral core-fucosylated mono- and digalactosylated structures, present at m/z values of 2056.1 and 2260.2, respectively, which we commonly observe to be associated with IgG (Alley, Vasseur, et al. 2012; Svoboda et al. 2012), were observed to be significantly decreased by ~ 40 – 50% in their abundance levels only when the control set consisting of samples provided by the former smokers was compared with the lung cancer samples provided by former smokers. The statistical testing of these groups for these two structures produced P -values which were both in the 10^{-3} range, whereas their AUC values were 0.21 and 0.176, respectively. These results are summarized in Table II.

The statistical analysis of the different sample cohorts, factoring in the smoking histories of their constituents, revealed some interesting trends. First, we were able to detect significant differences between the control groups of non-smokers and former smokers, indicating that the glycomic profiles of patients can be altered by a relatively short smoking history (9.7 years), even after having stopped. Interestingly, the group of control subjects who had previously smoked seemingly bridged the gap between the NSCLC patients who had smoked and control individuals who never smoked. Additionally, when comparing the control group of non-smokers to the NSCLC group of former smokers, we saw significantly changed P -values and AUC scores (data not shown), when compared with the previous analysis which did not separate the control group by their smoking history. Finally, there appeared to be a consistent decrease in both the P -values and AUC scores when comparing the control group of former smokers to the lung cancer cohort of former smokers (data not shown). These values saw significant declines and illustrate that the group most at risk for NSCLC, the former smokers, may be the most difficult samples in which to detect that condition. Although these values were still statistically significant, they were not nearly as robust as the values comparing the NSCLC group to the control group of non-smokers.

Exoglycosidase analysis of structures

The fucose residue appears to be an important substituent, as it has been shown to be elevated in a number of diseases (Gornik et al. 1999, 2007; Kyselova et al. 2007, 2008; Olewicz-Gawlik et al. 2007; Arnold et al. 2008; Gornik and Lauc 2008; Mehta and Block 2008; Bones et al. 2011; Kang et al. 2011; Alley, Vasseur, et al. 2012). Perhaps, just as important as the overall level of fucosylation is the fucose location on the glycan structure. Its possible locations are on either the core or an outer arm, where it may be present as one of the sialyl Lewis epitopes. An elevated presence of the genes encoding for the different $\alpha 1,3$ -fucosyltransferases in lung cancer patients have been associated with increased levels of sialyl Lewis epitopes,

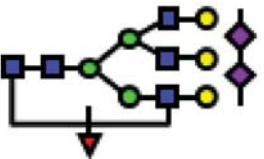
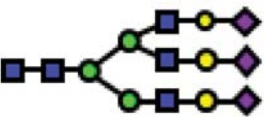
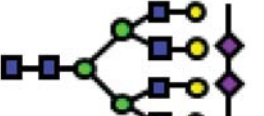
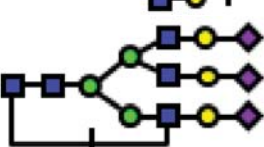
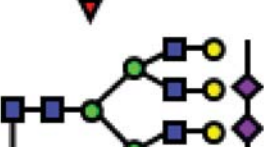
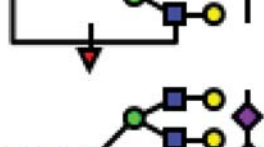
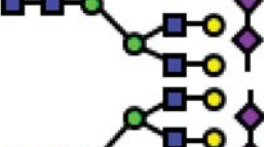
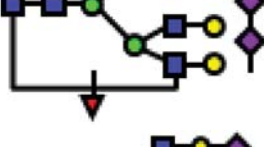
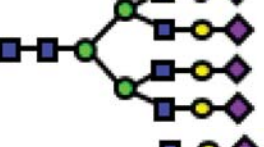
Table II. Average relative intensities for selected glycans for control samples provided by individuals who never smoked (control NS), control samples donated by former smokers (control FS) and samples given by small-cell lung carcinoma patients who were former smokers (NSCLC FS)

Structure	<i>m/z</i>	AVG REL INT			<i>P</i> -values		AUC values	
		Control NS	Control FS	NSCLC FS	Control NS vs control FS	Control FS vs NSCLC FS	Control NS vs control FS	Control FS vs NSCLC FS
	1852.0	7.79	7.94	6.90	0.898	0.445	0.529	0.357
	2056.1	9.21	8.15	5.41	0.250	3.46×10^{-3}	0.305	0.210
	2097.1	1.37	1.84	1.13	0.203	0.034	0.551	0.288
	2260.2	3.68	3.26	1.72	0.507	4.84×10^{-3}	0.283	0.176
	2301.2	1.88	1.80	1.42	0.587	0.144	0.433	0.238
	2447.2	8.63	7.74	8.64	0.121	0.223	0.342	0.624
	2505.3	0.719	0.656	0.547	0.354	0.073	0.422	0.292
	2621.3	3.43	3.30	2.84	0.716	0.230	0.353	0.386
	2808.4	36.3	38.2	42.1	0.287	0.025	0.690	0.717
	2896.5	0.599	0.579	0.576	0.683	0.966	0.444	0.541
	2982.5	3.33	3.84	4.20	0.087	0.316	0.690	0.574
	3070.6	0.182	0.181	0.197	0.502	0.772	0.551	0.489
	3227.6	1.11	0.87	1.24	0.185	0.053	0.348	0.677

Continued


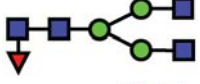
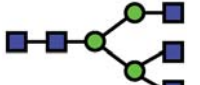



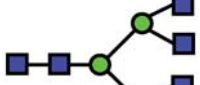

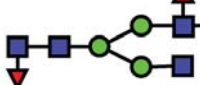


Downloaded from https://academic.oup.com/glycob/article/22/12/1684/1987969 by guest on 25 April 2024

Table II. (Continued)

Structure	<i>m/z</i>	AVG REL INT			<i>P</i> -values		AUC values	
		Control NS	Control FS	NSCLC FS	Control NS vs control FS	Control FS vs NSCLC FS	Control NS vs control FS	Control FS vs NSCLC FS
	3431.7	0.214	0.331	0.422	1.87×10^{-3}	0.040	0.813	0.665
	3618.8	4.29	3.50	4.80	0.129	0.058	0.348	0.724
	3706.9	0.238	0.203	0.246	0.384	0.151	0.454	0.633
	3792.9	1.04	2.15	3.10	1.20×10^{-4}	0.031	0.903	0.705
	3881.0	0.0404	0.0650	0.0933	0.006	0.030	0.797	0.724
	4068.0	0.173	0.134	0.186	0.180	0.058	0.358	0.683
	4242.1	0.045	0.081	0.131	1.78×10^{-3}	0.027	0.828	0.759
	4429.2	0.134	0.098	0.178	0.204	0.030	0.374	0.740
	4603.3	0.0386	0.0696	0.140	6.00×10^{-3}	0.031	0.786	0.800

The *P*-values, and AUC score for the statistical comparisons of control NS with control FS and control FS with NSCLC FS groups are also included.

Table III. Average relative intensities and *P*-values and AUC scores for the statistical comparisons of the different states of health for the products of the exoglycosidase analysis

Structure	<i>m/z</i>	Avg Rel Int			<i>P</i> -values		AUC scores	
		Control NS	Control FS	NSCLC NS	Control NS vs control FS	Control FS vs NSCLC FS	Control NS vs control FS	Control FS vs NSCLC FS
	1677.8	58.0	61.9	56.5	0.036	0.011	0.688	0.246
	1851.9	16.8	16.6	15.2	0.891	0.270	0.475	0.384
	1923.0	16.0	11.9	14.5	9.17×10^{-4}	0.014	0.163	0.76
	2056.0	0.358	0.718	0.981	3.13×10^{-4}	0.064	0.9	0.678
	2097.1	4.67	4.05	4.73	0.132	0.257	0.344	0.627
	2168.1	2.04	1.28	1.77	9.01×10^{-3}	0.045	0.190	0.704
	2230.1	0.098	0.059	0.089	0.074	0.056	0.398	0.695
	2301.2	1.56	2.91	4.99	2.32×10^{-5}	3.5×10^{-3}	0.928	0.781
	2342.2	0.039	0.027	0.047	0.037	7.97×10^{-3}	0.303	0.867
	2475.2	0.218	0.104	0.223	0.080	9.47×10^{-3}	0.402	0.811
	2546.3	0.177	0.324	0.765	2.09×10^{-4}	1.27×10^{-3}	0.868	0.814

which, in turn, have been linked to a poorer patient prognosis (Ogawa et al. 1996). To gain a more detailed understanding of the locations where the fucose units reside, and the changes in their abundances within the different sample

sets, the previously released glycan samples were subjected to an exoglycosidase digestion using a non-specific sialidase and a β 1-4,6-galactosidase. During the course of this digestion, if the fucose monosaccharides were located on

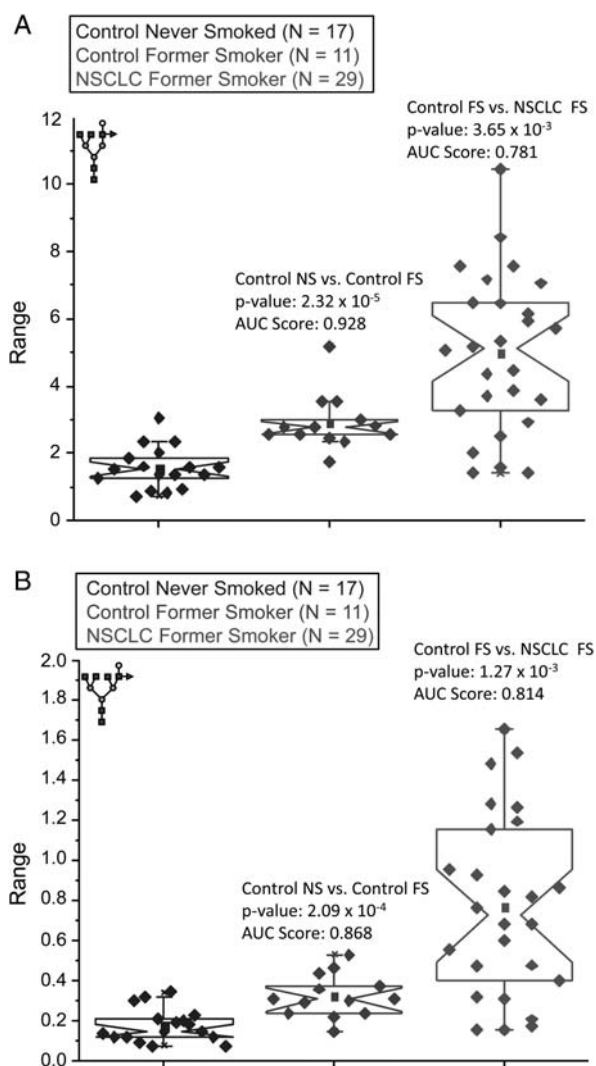


Fig. 6. Notched-box plots depicting increased levels of outer-arm fucosylation in (A) a tri-antennary structure and (B) a tetra-branched product of an exoglycosidase digestion. The abundances of these types of analytes increased from the control sample provided by individuals who had never smoked, to those obtained from former smokers and to NSCLC samples given by former smokers.

an outer arm, the activity of this β -galactosidase becomes restricted, leaving the galactose attached to its GlcNAc residue. Therefore, two isomeric structures prior to the exoglycosidase digestion that differed only in the location of a fucose will now be present with different molecular masses as a result of the digestion procedure. Such a mass difference can be detected by MALDI MS, permitting any changes in the ratios of core- to outer-arm fucosylation to be quantified.

The MALDI mass spectra of the permethylated structures following the exoglycosidase digestion contained a total of 12 distinct m/z values, which could be correlated with the products of the digestion procedure. The α -fucosylated bi-, tri- and tetra-antennary products were observed, along with their three core- and outer-arm fucosylated analogs. Additional signals that we observed were determined to be tri- and tetra-antennary

doubly fucosylated analytes, which were present as either a core- and-outer-arm doubly fucosylated structures or doubly outer-arm fucosylated analytes.

In this analysis, the results of which are summarized in Table III, each of the bi-, tri- and tetra-antennary structures possessing an outer-arm fucose was found to be nearly doubled in their abundance levels when the control samples provided by non-smokers were compared with those of former smokers. These elevated levels were statistically supported with P -values in the 10^{-4} – 10^{-5} range and AUC scores that were all above 0.85. These three structures were further elevated in their abundances in the lung cancer samples provided by the former smokers, with the most significant increases occurring for the tri- and tetra-branched analytes, as shown by the notched-box plots presented as Figure 6A and B, for the tri- and tetra-antennary structures, respectively. The data for both of these structures resulted in P -values in the 10^{-3} range with AUC scores of 0.78 and 0.81, respectively.

Sialic acid linkage analysis of structures

To investigate the possible alterations in the ratios of the sialic acid linkages of the individual glycan structures, the samples were subjected to our differential amidation analytical procedure (Alley and Novotny 2010). In this procedure, those sialic acids which are linked in an $\alpha 2,3$ manner form internal esters, or lactones, preventing their amidation. This lactonization is not possible for $\alpha 2,6$ -linked sialic acids, allowing them to become amidated. During the subsequent permethylation, the ester linkage is cleaved and the resulting carboxylate accepts a methyl group, whereas the nitrogen of the amide becomes doubly permethylated, causing a 13-Da mass difference in the MALDI mass spectra.

The linkage isomers of the fucosylated tri-sialylated tri-antennary carbohydrate produced an interesting trend. The notched-box plot depicting their levels associated with this glycan is shown in Figure 7. Although the exact location of the fucose unit remained unclear in this analysis, there seemed to be a statistically supported decrease in the levels of isomers possessing only $\alpha 2,3$ -linked sialic acids between the control samples provided by individuals with no smoking history and those who were former smokers. This comparison produced a P -value of 0.0413 and an AUC score of 0.265. The suppression of this isomer was further accentuated in the lung cancer samples provided by former smokers, with a P -value of 2.14×10^{-3} and an AUC score of 0.161 being calculated for the comparison of the different states of health for the former smokers. The structure containing two $\alpha 2,3$ -linked sialic acids also appeared to be decreased in both sample sets comprised of former smokers, when compared with the control cohort of individuals with no smoking history, with this particular structure accounting for $\sim 15\%$ of the signal intensity for this group, and $\sim 12\%$ in both sample sets provided by the former smokers. However, statistical testing provided inconclusive results. Conversely, the isomer containing two $\alpha 2,6$ -linked sialic acids was noticeably increased in its abundance levels for both groups of former smokers. This particular isomer contributed ~ 81 and 83% of the total intensity associated with this particular glycan for the control samples provided by

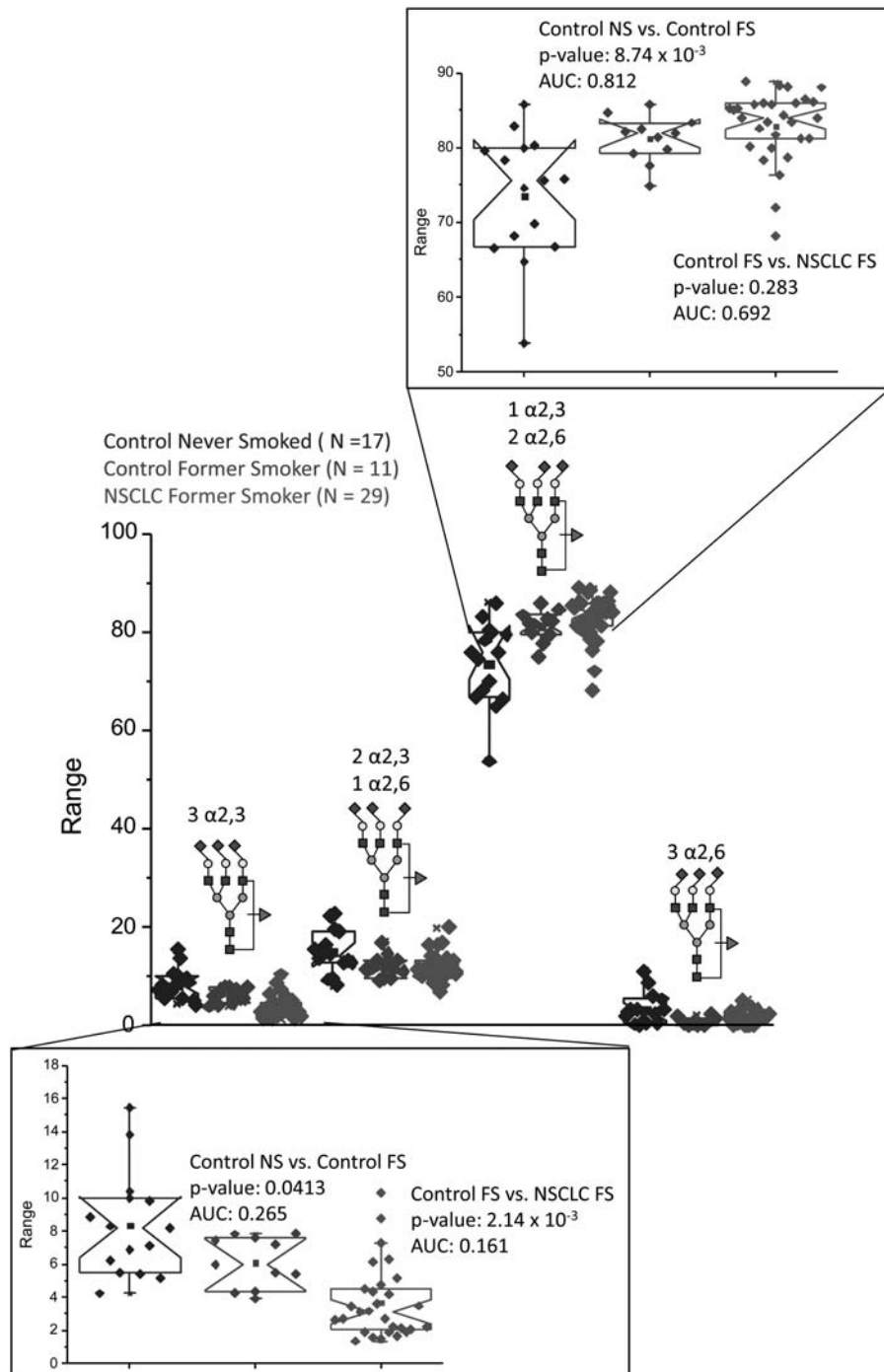


Fig. 7. Notched-box plots depicting the results of an amidation/lactonization analysis to determine the ratios of the different sialic acid linkages. The most significant differences were observed for the isomer composed of three $\alpha 2,3$ -linked sialic acids and the isomer possessing two $\alpha 2,6$ -linked sialic acids.

former smokers and the lung cancer samples provided by former smokers, respectively. In the control samples acquired from the individuals with no smoking history, this isomer produced $\sim 73\%$ of the total signal associated with this glycan. The elevated level of this isomer passed the statistical criteria when the different control sample sets were compared; a P -value of 8.74×10^{-3} was calculated and the AUC score

was 0.812. However, a comparison of the groups composed of former smokers resulted in a P -value significantly above 0.05. The isomer composed of only $\alpha 2,6$ -linked sialic acids that was present in many samples across all of the sample sets at fairly low abundances produced a wide spread in the data. Therefore, it is difficult to extract meaningful conclusions for this isomer.

Table IV. Average relative intensities, *P*-values, standard error of the mean (SEM values) and AUC scores for the comparisons of the different states of health based on the amidation/lactonization data to examine differences in sialic acid linkages

Structure	<i>m/z</i>	# of α 2,3	# of α 2,6	Control NS		Control FS		NSCLC FS		Control NS vs Control FS		Control FS vs NSCLC FS	
				Avg Rel Int	SEM	Avg Rel Int	SEM	Avg Rel Int	SEM	<i>P</i> -value	AUC score	<i>P</i> -value	AUC score
	2447.2	1	0	18.3	10.7	17.1	5.0	18.7	10.6	0.08	N/A	0.02	0.74
	2460.2	0	1	81.7	2.4	82.9	1.0	81.3	2.4	0.08	N/A	0.02	0.26
	2621.3	1	0	21.4	11.7	22.4	11.8	22.2	17.6	342	N/A	0.91	N/A
	2634.4	0	1	78.6	3.2	77.6	3.4	77.8	5.0	0.342	N/A	0.91	N/A
	2808.5	2	0	32.4	7.9	33.7	11.4	34.3	9.4	0.335	N/A	0.59	N/A
	2821.5	1	1	34.2	9.8	33.7	8.4	33.5	7.5	725	N/A	0.81	N/A
	2834.5	0	2	33.4	11.5	32.6	8.6	32.2	10.6	0.567	N/A	0.71	N/A
	2896.5	1	0	34.1	13.6	29.7	11.5	32.6	20.4	0.015	0.212	0.18	N/A
	2909.5	0	1	65.9	7.0	70.3	4.9	67.4	9.9	0.015	0.788	0.18	N/A
	2982.5	2	0	7.1	31.3	7.8	29	2.9	68.1	0.447	N/A	5.70×10^{-8}	0.059
	2995.5	1	1	22.9	20.8	28.1	18.3	24.9	24.5	0.013	0.77	0.13	N/A
	3008.6	0	2	70.1	9.4	64.1	10.7	72.2	9.6	0.035	0.273	2.00×10^{-3}	0.79
	3070.6	1	0	41.1	13.6	40.4	17.0	39.8	17.8	0.773	N/A	0.82	N/A
	3083.6	0	1	58.9	9.5	59.6	11.5	60.2	11.7	0.773	N/A	0.82	N/A
	3227.6	2	0	13.9	32.8	17.8	42.9	14.8	47.4	0.11	N/A	0.24	N/A
	3240.6	1	1	28.6	11.7	30.6	13.0	29.2	18.3	0.16	N/A	0.44	N/A
	3253.6	0	2	57.6	12.9	51.6	21.7	56	21.2	0.11	N/A	0.3	N/A
	3257.7	2	0	18.2	24.4	18	30.3	17.0	33.4	0.91	N/A	0.63	N/A
	3270.7	1	1	47.3	10.4	45.1	9.1	45.7	15.4	0.23	N/A	0.8	N/A
	3283.7	0	2	34.5	15.9	36.9	12.4	37.3	19.8	0.24	N/A	0.88	N/A
	3431.7	2	0	10.8	25.5	10.0	19.2	9.1	23.2	0.38	N/A	0.247	N/A
	3444.9	1	1	60.4	13.2	66.5	8.5	63.7	9.1	0.04	0.733	0.19	N/A
	3457.9	0	2	28.8	28.4	23.6	20.5	27.2	24.9	0.07	N/A	0.12	N/A
	3618.8	3	0	2.2	44.2	2.7	34.3	2.3	43.5	0.21	N/A	0.3	N/A
	3631.9	2	1	10.2	31.8	8.6	16.3	9.8	29.2	0.13	N/A	0.17	N/A
	3644.9	1	2	70.3	5.3	66.1	4.9	61	10.6	6.8×10^{-3}	0.18	0.016	0.273
	3657.9	0	3	17.2	24.7	22.6	15.2	26.9	31.3	2.3×10^{-3}	0.83	0.112	N/A
	3792.9	3	0	8.2	38.7	6.3	27.4	3.7	60.9	0.0754	N/A	1.00×10^{-3}	0.16
	3805.9	2	1	14.8	30.6	12.2	19.6	11.7	25.0	0.1	N/A	0.61	N/A
	3818.9	1	2	74.0	11.8	80.6	4.6	83	5.9	0.027	0.7	0.138	N/A
	3831.9	0	3	3.0	109.5	1.0	98.9	1.6	74.3	0.06	N/A	0.15	N/A

Continued

Table IV. (Continued)

Structure	<i>m/z</i>	# of α 2,3	# of α 2,6	Control NS		Control FS		NSCLC FS		Control NS vs Control FS		Control FS vs NSCLC FS	
				Avg Rel Int	SEM	Avg Rel Int	SEM	Avg Rel Int	SEM	<i>P</i> -value	AUC score	<i>P</i> -value	AUC score
	4068.0	3	0	18.3	25.1	19.2	15.2	17.6	25.0	0.58	N/A	0.26	N/A
	4081.0	2	1	36.0	12.5	30.7	20	35.8	23.0	0.02	0.206	0.07	N/A
	4094.0	1	2	33.6	20.7	34.5	16.7	35.6	19.0	0.73	N/A	0.65	N/A
	4107.0	0	3	12.0	34.5	15.6	33.7	11.1	33.0	0.07	N/A	4.00×10^{-3}	266
	4242.1	3	0	24.1	18.1	21.7	21.1	17.6	30.9	0.19	N/A	0.03	27
	4255.1	2	1	30.7	14.3	32.6	15.1	39.2	15.8	0.32	N/A	3.00×10^{-3}	0.79
	4268.1	1	2	25.5	16.2	27.0	17.9	31.6	20.6	0.42	N/A	0.039	0.73
	4281.1	0	3	19.6	10.8	18.7	22.7	11.7	50.8	0.47	N/A	9.70×10^{-4}	0.18
	4429.2	4	0	16.7	25.3	19.3	16.6	19.1	30.5	0.11	N/A	0.92	N/A
	4442.2	3	1	20.6	14.8	21.9	9.3	23.4	16.5	0.24	N/A	0.24	N/A
	4455.2	2	2	30.4	18.8	25.4	19.6	29.2	24.7	0.03	0.03	0.11	N/A
	4468.2	1	3	18.8	15.5	18.3	10.9	17.4	21.4	0.61	N/A	0.45	N/A
	4481.2	0	4	13.4	22.4	15.1	25.1	10.9	39.5	0.21	N/A	6.00×10^{-3}	0.23
	4603.3	4	0	20.3	16.4	20.7	22.8	17.5	18.0	0.82	N/A	0.019	0.31
	4616.3	3	1	21.8	9.8	20.9	17	21.2	15.6	0.41	N/A	0.77	N/A
	4629.3	2	2	22.5	16.8	23.2	17.2	32.1	25.1	0.66	N/A	1.00×10^{-3}	0.82
	4642.3	1	3	17.9	20.9	18.4	25.2	15.7	29.0	0.74	N/A	0.11	N/A
	4655.3	0	4	17.5	23.4	16.9	23.9	13.5	27.1	0.69	N/A	0.014	0.31

Control NS, control non-smoker; control FS: control former smoker; NSCLC FS, non-small cell lung carcinoma patient (former smoker).

Interestingly, the isomers of the α -fucosylated tri-sialylated tri-antennary structure containing three and two α 2,3-linked sialic acids did not appear to be significantly different in any of the sample groups. However, the α -fucosylated analog indicated that the levels of the isomer possessing two α 2,6-associated sialic acids were decreased in their abundance levels. This isomer, which accounts for just over 70% of total ion intensity for this glycan, was observed for the control samples with no smoking history. An intermediate value of 66% was noted for the control samples given by former smokers, and the lowest contribution from this isomer at 61% being calculated for the lung cancer samples was yielded by former smokers. These decreased abundance levels were statistically significant, based on a *P*-value of 6.8×10^{-3} and an AUC score of 0.180 when the two types of control samples were compared. The comparison of the control group and lung cancer groups of former smokers also appeared to be statistically relevant, with a *P*-value of 0.016 and the AUC score determined as 0.273. The overall results for this analysis are given in Table IV.

Discussion

The current study continues our series of serum-based glycomic investigations of different cancers to identify potential

markers of specific malignancies and to begin an assessment of the specificity of the glycomic alterations. Using only a few microliters of serum, we have performed a comprehensive N-linked glycomic profiling, a more detailed analysis of the location of the fucose residues and an in-depth analysis of the variations of sialic acid linkages in the different states of health associated with smoking and lung cancer. The clinical samples were provided mainly by the individuals with no smoking history who acted as controls, individuals who were former smokers, presented as a “second control set”, and patients who were diagnosed with lung cancer who had smoked previously. Indeed, this study is among the first to examine the potential effects of smoking on the glycomic profiles (Knezevic et al. 2010; Arnold et al. 2011), but undoubtedly more detailed than the previous investigations.

Often, smoking is associated with chronic obstructive pulmonary disease, a condition that is generally not observed in the non-smoking population. In the current literature, relatively little is known about the effects of this condition on the types of N-linked structures, such as those that we analyzed in this study. However, chronic inflammation has been reported to result in an overproduction of mucins (Hauber et al. 2006; Voynow et al. 2006), molecules typically containing high levels of O-linked glycans, whereas alterations to the

carbohydrate patterns of mucins seem to be also altered in cystic fibrosis (Schulz et al. 2005). If properly designed, studies of this condition concerning serum glycomic profiles could prove interesting.

In this study, the data indicated quantitative alterations to 29 distinct ions corresponding to the generally accepted glycan structures, when disregarding the smoking status of the control and lung cancer groups. Among the glycans that appeared to be the most significantly altered were the neutral galactosylated structures, a trend which we, and another group, previously had observed for ovarian cancer (Saldova et al. 2007; Alley, Vasseur, et al. 2012). The abundance levels of these glycans seemed to be correlated with each cohort group in the present study, with its highest levels occurring in the control group with no smoking history, an intermediate value for the control samples who were former smokers and its lowest levels being recorded for the lung cancer patients who had formerly smoked. According to our previous results (Alley, Vasseur, et al. 2012; Svoboda et al. 2012), these types of structures are mainly associated with IgG, whereas the galactosylated structures appear to be important in the mannose-binding lectin mechanism to activate the complement factor response (Malhotra et al. 1995; Rudd et al. 1995). In our previous studies, the α -galactosylated structures associated with IgG were generally increased in their intensity levels. In particular, this seemed evident for the bi-antennary, core-fucosylated structure appearing at an m/z value of 1852.0 (Kyselova et al. 2007, 2008; Kang et al. 2011; Alley, Vasseur, et al. 2012). However, in this study, this structure was present with approximately the same abundance levels in each of the sample groups. It may be possible that the fairly constant levels of this structure could be associated with lung cancer.

Our past studies (Kyselova et al. 2007, 2008; Kang et al. 2011; Alley, Vasseur, et al. 2012) have also suggested that the tri- and tetra-antennary structures with varying degrees of sialylation and fucosylation may be key glycans to distinguish control samples from those of a specific cancer under investigation. This trend was also observed in this study, with the changes associated with the fucosylated analogs generally being the more statistically supported alterations. Similar to the several of our past studies (Kyselova et al. 2007, 2008; Kang et al. 2011; Alley, Vasseur, et al. 2012), a fucosylated tri-antennary tri-sialylated structure was markedly elevated in the control samples provided by individuals who had formerly smoked and continued to be increased in its levels in the lung cancer cohort of former smokers. Interestingly, this particular glycan has been the one observed as increased in breast cancer cell lines (Lattova et al. 2010), and following a treatment with antibody drugs, its abundance levels were suppressed. In combination, these results show that this particular structure may lack the specificity to accurately “diagnose” a particular cancer, but it may find its value in monitoring disease progression/regression. To the best of our knowledge, the serum protein(s) to which this oligosaccharide is attached has yet to be identified, but the protein itself may also be of interest to study in more detail.

The altered elevation of the fucosylated tri-antennary tri-sialylated oligosaccharide highlights two of the main goals of our cancer-related studies: (i) to propose a set of glycomic

changes between the pathological samples and those which were provided by “control” individuals and (ii) to evaluate and compare the changes observed in the data sets to our previous studies to facilitate the understanding of the specificity of the alterations for a particular cancer. Supplementary data, Table SIII summarizes the glycans which we have suggested as indicators for our previously published cancer studies, including the breast (Kyselova et al. 2008; Alley et al. 2010), liver (Goldman et al. 2009; Kang et al. 2011), esophageal (Mechref et al. 2009), prostate (Kyselova et al. 2007) and ovarian (Alley, Vasseur, et al. 2012), along with the results of this study. A few oligosaccharides which were differentially expressed on the serum glycoproteins of lung cancer patients were common to several (three or more) of the other cancers, which we had studied in the past and include several structures that we believe are attached primarily to IgG (Alley, Vasseur, et al. 2012; Svoboda et al. 2012). These types of analytes were observed at m/z values of 2056.1, 2260.1 and 2301.2 (see Supplementary data, Table SIII, for their structures). Additionally, some blood-serum glycoprotein carbohydrates with altered abundance levels in the lung cancer patients are generally associated with the acute-phase proteins that were also common in at least three other cancers. Such glycans were recorded at m/z values of 2621.3, 2808.4, 3431.7, 3881.0 and 4242.1. Interestingly, all of these oligosaccharides, except for that observed at the m/z value of 2808.4, are fucosylated. Thus, it seems likely that while increased levels of fucosylation may indicate the presence of a cancer and could possibly be correlated with its stage (Alley, Vasseur, et al. 2012), fucosylation alone probably lacks the specificity as a “biomarker” for a particular condition.

Nevertheless, a number of carbohydrates with altered levels in lung cancer samples appeared to be more specific for this condition and overlapped with no more than two of the other cancers that we have previously studied. Some of these structures are most likely associated with IgG, having been detected at m/z values of 1595.8, 1636.8, 1799.9, 1881.9, 2003.9, 2026.0, 2330.9, 2505.5, 2662.3, 3227.6 and 3401.6, indicating the potential value of specifically studying the glycosylation of the selected immunoglobulins. Other carbohydrates with altered expressions in lung cancer, which were also noted in not more than two other cases, included ions detected at m/z values of 2417.2, 2488.2, 2982.5, 3618.8, 3792.9, 3967.0, 4429.2 and 4603.3. These are often associated with the acute-phase responders (Arnold et al. 2008; Gornik and Lauc 2008).

In an attempt to improve the measurement specificity, we have recently begun to perform more in-depth studies by examining the different isomeric structures. Similar to the breast (unpublished results) and ovarian cancers (Alley, Vasseur, et al. 2012), the glycans derived from blood serum glycoproteins of lung cancer patients showed an increase in their levels of outer-arm fucosylation of the di-, tri- and tetra-antennary structures. Conversely, the liver cancer favors an increased abundance of core-fucosylated glycans (Alley et al. in preparation), indicating that a certain level of specificity could be associated with the location of the fucose monosaccharide. The increased expressions of outer-arm fucosylation in our lung cancer study correlates well with a previous study

of lung cancer using a different analytical methodology (Arnold et al. 2011) and may be associated with the sialyl Lewis structures that are often associated with chronic inflammation and late-stage cancers (Arnold et al. 2008).

Additionally, we have developed a microscale differential amidation procedure that allows us to monitor the changes in the ratios of α 2,3- and α 2,6-linked sialic acids as a function of a pathological condition (Alley and Novotny 2010). This particular type of analysis seems to show promise by providing more specific structural information. An example of the capabilities of this method is demonstrated for the ion detected at an m/z value of 3792.9, a carbohydrate which was observed to be increased in its relative abundance in the breast (Kyselova et al. 2008; Alley et al. 2010), prostate (Kyselova et al. 2007), ovarian (Alley, Vasseur, et al. 2012) and lung cancers, thus limiting its ability to function as a specific marker. However, following the amidation procedure, no changes to the linkage ratios for this particular structure were observed in the ovarian cancer samples, whereas in serum samples provided by prostate cancer patients, this glycan featured increased levels of α 2,3-linked sialic acids for those isomers with three and two α 2,3-linked structures and lowered abundance levels for the isomers containing two and three α 2,6-linked sialic acids (Alley, Vasseur, et al. 2012). Interestingly, in the lung cancer samples analyzed in this study, the opposite trend was observed: the isomers composed of three and two α 2,3-linked sialic acids were decreased in their abundance levels, whereas that consisting of two α 2,6-attached sialic acids was actually increased. Although this analysis needs to be performed on other cancer sets, it seems plausible that by performing more detailed structural characterizations, and considering all of the observed changes in concert, the specificity of the glycomic measurements could be substantially improved.

One of the main objectives of this work was to examine the effects of smoking on the glycomic profiles and to provide a more detailed view complementing the previous studies that investigated the effects tobacco smoke and its exposure on serum proteins (Weiss et al. 1981). Several groups have earlier indicated that the exposure to tobacco smoke may induce increased expressions of the acute-phase proteins, including haptoglobin (Bradley et al. 1977) and ceruloplasmin (Davidoff et al. 1978). Elevated levels of complement 3c and α 1-acid glycoprotein (AGP) have also been reported for children exposed to environmental tobacco smoke (Shima and Adachi 1996), whereas the immune responses of the users of tobacco products may be suppressed (Moszczynski et al. 2001). However, little work has been done to study the effects of tobacco smoke on the blood serum glycome. One such study, which examined the serum protein-derived glycans for a very large number of individuals, found the effects of smoking on blood serum glycosylation to be quite significant (Knezevic et al. 2009). In fact, the authors of this study viewed these changes associated with smoking as being so important that they concluded that smoking was the single most important environmental parameter that may induce changes in the serum glycome. Indeed, based on this study, this group has suggested that the effects of smoking need to be considered and controlled in future glycomic studies.

Recently, a glycomic study of the sera of lung cancer patients was published using alternative analytical methods (Arnold et al. 2011). This work concluded that tri-sialylated structures were elevated in the comprehensive *N*-linked glycomes of the pathological samples and could be statistically correlated with the disease stage. When examining in more detail the glycomic patterns associated with haptoglobin, it appeared that increased levels of tri-sialylated carbohydrates and those possessing sialyl Lewis x structures were elevated in glycomic recordings of stage IV lung cancer patients when compared with control samples, though the full significance of these alterations could not be fully statistically confirmed. Additionally, the possibility exists for decreased abundances of certain haptoglobin-derived carbohydrates from lung cancer patients, as based on the suppressed intensities of two liquid-chromatographic peaks in that study. Unfortunately, several glycans could be present in each chromatographic peak and the key carbohydrates contributing to the lowered peak areas could not be unequivocally determined.

Given the dynamic concentration range of serum proteins, with several glycoproteins present with concentrations in the mg/mL range and the so-called “biomarkers” associated with tissue-leakage proteins expected to have pg/mL concentrations (Anderson and Anderson 2002), we believe that the changes we report here are primarily due to an acute-phase response. This biological reaction, which has been shown to be associated with late-stage cancer and inflammation (Arnold et al. 2008), in general, causes several highly abundant glycoproteins synthesized by the liver to be changed either up or down in their concentrations in blood. Several of these proteins have been studied to determine either their diagnostic or prognostic abilities for NSCLC. A key protein that has been reported to be altered in many different cancers is AGP (Shiyan and Bovin 1997; Van Den Heuvel et al. 2000; Lacunza et al. 2007; Imre et al. 2008). Although its reliability to accurately diagnose a particular cancer, as based on its concentration alone, appears to be questionable, its value as a predictor of the success of a treatment with docetaxel for NSCLC patients has been demonstrated (Bruno et al. 2003). In a study of 189 lung cancer patients, the individuals with higher levels of AGP had lower effective responses to the drug and shorter survival times, by an average of 10 months, than those with lower AGP levels (Bruno et al. 2003). A further analysis of the specific carbohydrates attached to this protein in the serum of individuals diagnosed with lung cancer (Hashimoto et al. 2004) indicated that those patients with elevated levels of fucosylated tri- and tetra-antennary structures for more than 2 weeks following surgery tended to have a poorer prognosis.

Haptoglobin and its glycosylation patterns have also been examined in several cancers, including lung cancer. One recent report has suggested that fucosylated haptoglobin is elevated in the serum of lung cancer patients (Tsai et al. 2011), similar to the patients suffering from pancreatic cancer (Lin et al. 2011). However, in the lung cancer patients, it was reported that a fucosylated tri-sialylated tri-antennary carbohydrate was elevated (Tsai et al. 2011). Based on lectin staining and tandem MS fragmentation, it was suggested that this glycan was enriched with α 2,6-linked sialic acids, which is in good agreement with the data that we present in this

communication. Interestingly, fucosylation has also been suggested as being capable of differentiating small cell vs non-small lung carcinomas (Kossowska et al. 2005). Proteins present in a 29-kDa gel-electrophoretic band tended to carry more fucose units than those in a 42-kDa band, where AGP and haptoglobin were identified. These authors have tentatively suggested that the fucose levels could be associated with patient survival (Kossowska et al. 2005). Enzyme-linked immunosorbent assay tests have also been employed to investigate haptoglobin and its glycoforms, indicating that haptoglobin is elevated in lung cancer patients along with its levels of sialylation and fucosylation (Hoagland et al. 2007).

The glycosylation of these types of proteins is apparently influenced by a particular pathological condition and could be the result of signaling molecules (i.e. cytokines) released by the tumors. Upon reaching the liver, these molecules seem to alter the syntheses of the oligosaccharides, as suggested by the report that indicated that interleukins (IL) 1 and 6 resulted in enhanced levels of fucosylation for AGP produced by human hepatocytes (Azuma et al. 2000). Lung cancers are known to release several cytokines, including IL-6 and IL-8 (Takizawa et al. 1993; Alleva et al. 1994), with increased levels of IL-6 being associated with a poor prognosis for both African-American and Caucasian populations (Enewold et al. 2009). Interestingly, when healthy human bronchial mucosa fragments were incubated with these types of molecules, increased levels of the genes encoding for the α 1,3/4-fucosyltransferases and the α 2,3- and α 2,6-sialyltransferases, which could result in increased abundances of sialyl Lewis epitopes, were observed (Delmotte et al. 2002). Additionally, stimulation of a myeloma cell line with IL-6 induced the increased levels of GlcNAc transferases IV and V (Nakao et al. 1990), the enzymes responsible for catalyzing the formation of the tri- and tetra-antennary structures, respectively, whereas GlcNAc transferase V was also reported as elevated in liver cancer tissues (Yao et al. 1998). Consequently, it seems plausible that the many of the key glycosyltransferases (fucosyl, sialyl and GlcNAc) expressed in the liver are sensitive toward stimulation by the various cytokines and could be responsible for the alterations that we have observed here for the lung and, additionally, other cancers studied previously in this group.

Conclusions

We report here a comprehensive glycomic analysis of human blood serum, which includes our total N-linked analysis of permethylated glycans, an exoglycosidase digestion to pinpoint the location of the fucose residues and an examination of the sialic acid linkages to determine their potential in distinguishing lung cancer samples from control sets. Further analyses were conducted to evaluate the effects of smoking on the serum glycomic patterns. The results reveal significant trends associated with the disease samples, including an increase in the abundance levels of the fucosylated tri- and tetra-antennary structures and elevations in outer-arm fucosylation, along with changes in the overall pattern of sialic acid linkages. Most prominently, a fucosylated tri-sialylated tri-antennary structure indicated decreased abundance levels of the isomers containing three and two α 2,3-linked sialic acids

and elevated levels of an isomer with two α 2,6-associated sialic acids in the control samples provided by former smokers and former smokers diagnosed with lung cancer. In many cases, for each of the glycomic tests that we performed, the control samples provided by the former smokers were present at intermediate values, in-between the control samples obtained from the control individuals with no smoking history and the lung cancer patients who were former smokers. These results suggest that those individuals most at risk for developing lung cancer, the former smokers, may be difficult to diagnose using these types of investigations. Furthermore, our data also suggest that the smoking histories of subjects enrolled in such studies need to be considered.

Materials and methods

Materials

Trifluoroacetic acid (TFAA) was purchased from EMD Chemicals (Gibbstown, NJ), whereas chloroform, acetic acid, methanol, acetonitrile (ACN) and *N,N*-dimethylformamide (DMF) were received from Mallinckrodt Baker (Phillipsburg, NJ). Peptide-*N*-glycosidase F (PNGase F), isolated from *Escherichia coli*, was purchased from Cape Cod Company (East Falmouth, MA), whereas sialidase A, extracted from a recombinant gene from *Arthrobacter ureafaciens* expressed in *E. coli*, and β -(1-3,4,6)-galactosidase, isolated from bovine testes, were from ProZyme, Inc. (Hayward, CA). Nonidet P40 was a product of Roche (Indianapolis, IN) and sodium dodecyl sulfate (SDS) was received from Bio-Rad Laboratories (Hercules, CA). The MALDI matrix, 2,5-dihydroxybenzoic acid (DHB), was purchased from Alfa Aesar (Ward Hill, MA). All micro-spin columns (C₁₈, graphite, amine and “empty”) were purchased from Harvard Apparatus (Holliston, MA). Additional reagents (β -mercaptoethanol, sodium hydroxide beads and methyl iodide) used in sample treatment and those needed to synthesize 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM) were received from Sigma-Aldrich (St Louis, MO).

Blood serum samples

Blood serum samples were provided by the Hoosier Oncology Group, Inc. (Indianapolis, IN) and collected according to its banking protocol BNK09-138. Approximately 10 mL of blood was drawn into glass red top tubes. After allowing the samples to coagulate for 30 min, they were centrifuged at 1.2 kg for 30 min. Within 1 h of being drawn, the serum was divided equally into three labeled cryovials and then immediately frozen at -70°C until time of shipment. Serum samples needed to be submitted prior to the patient receiving any treatment for lung cancer and were thus considered to be baseline samples.

PNGase F digestion

A 10- μL aliquot of each serum sample was first diluted with 35 μL of buffer composed of 10 mM sodium phosphate (pH 7.5), 0.1%/0.1% β -mercaptoethanol/SDS. The samples were simultaneously denatured and their protein disulfide bonds were reduced by incubating the samples at 60°C for 30 min. After allowing the samples to cool to the ambient temperature, a 5- μL aliquot of 10% Nonidet P40 was added and the

samples were further equilibrated at room temperature for an additional 20-min period to ensure that the SDS molecules had sufficiently partitioned into the Nonidet P40 micelles. PNGase F (equivalent to 1 IU) was then added to release the glycans from their protein backbones, whereas the samples were incubated for 18 h at 37°C.

Purification of glycans through C₁₈ and graphite beds

The enzymatically released glycans were purified in a two-step procedure that first employed C₁₈ micro-spin columns to remove proteins and other relatively hydrophobic molecules present in the reaction media. Subsequently, graphite micro-spin columns were employed to remove salts and other small-molecule contaminants. The C₁₈ columns were first conditioned with three 400- μ L aliquots of a solution composed of 85%/15%/0.1% ACN/water/TFAA (buffer B) and then re-equilibrated with three 400- μ L aliquots of a solution composed of 5%/95%/0.1% ACN/water/TFAA (buffer A). The samples were then loaded onto the medium, centrifuged, reapplied a second time and centrifuged again. The stationary phase was then washed twice with 200- μ L aliquots of buffer A to ensure a high recovery of non-retained glycans from the solid-phase extraction medium. The graphite columns were conditioned and re-equilibrated using the same procedure as employed for the C₁₈ columns. For this medium, the samples were loaded onto the columns and centrifuged twice. Salts and other contaminants were removed by washing the graphite medium twice with 200- μ L aliquots of buffer A. The glycans were then eluted with two 200- μ L aliquots of buffer C (25%/75%/0.1% ACN/water/TFAA). Since the exoglycosidase digestion procedure (described later) generates several neutral structures, some of which are already present on serum glycoproteins, the samples were re-digested and the neutral carbohydrates were eluted from the graphite columns through the application of 200 μ L of a solution composed of 25%/75% ACN/water. The sialylated structures were then eluted as described previously. Finally, the samples were dried completely in a vacuum centrifuge.

Reduction of glycans to alditols

Following their solid-phase purification, the glycans were reduced to their alditol forms using our previously described protocol (Huang et al. 2002; Alley et al. 2010). Briefly, this involved the addition of a 10- μ L aliquot of a 10-mg/mL aqueous solution of ammonia–borane complex and incubating the samples at 60°C for 1 h. After allowing the samples to cool to the room temperature, 20 μ L of a 10% acetic acid solution were added twice over a 2-h time period to eliminate any unreacted ammonia–borane complex. The samples were then dried under vacuum. Finally, three 100- μ L aliquots of methanol were added to the samples to remove borane salts as their volatile methyl esters. After drying, the samples were reconstituted in a 10- μ L aliquot of water and divided into four 2.5- μ L aliquots for further analyses.

Exoglycosidase digestion

In order to pinpoint the locations of fucose residues on a glycan structure, one set of aliquots was digested with a

non-specific sialidase (1.0 U) to remove all sialic acid residues and, subsequently, with \sim 10 U of a β (1-3,4,6) galactosidase. The pH value for all exoglycosidase digestion solutions was adjusted to 5.5 through the addition of a 10- μ L aliquot of 20 mM sodium acetate buffer. These digestions were performed at 37°C for 18 h, after which the samples were subsequently dried, resuspended in a 5- μ L aliquot of water and permethylated, as described in the section of this communication discussing the specific details of this derivatization procedure.

Amidation/lactonization of sialic acids

The potential differences in the ratios of the sialic acid linkages, conjugated as either α 2,3 or α 2,6, associated with the pathological samples were investigated using our amidation/lactonization procedure (Alley and Novotny 2010). Briefly, the samples were resuspended in 10 μ L of a 0.5-M ammonium chloride solution (pH 6.5) and incubated for 2 h at 50°C to ensure the complete lactonization of the α 2,3-linked structures. Subsequently, 10 μ L of a freshly prepared solution of 0.5 M DMT-MM (synthesized according to the previously published procedure; Kunishima et al. 1999) was added to the samples prior to their incubation at 40°C for 36 h. Following the reaction, the samples were diluted to 140 μ L with ACN and then purified using amino micro-spin columns. Prior to the sample application, the columns were conditioned with three 400- μ L aliquots of a solution composed of 5%/95%/0.1% water/ACN/TFAA (solution A) and then re-equilibrated with a solution consisting of 85%/15%/0.1% ACN/water/TFAA (solution B). The samples were recycled through the media a total of three times and the packings were washed with two 200- μ L aliquots of solution B. The glycans were then eluted with two 200- μ L aliquots of solution A and dried under vacuum. They were stored at -20°C until the time of their permethylation.

Permethylation

The samples were permethylated using our “static” approach to permethylation (Alley et al. 2010), which does not utilize flow during the reaction (Kang et al. 2005). First, permethylation reactors were constructed by suspending sodium hydroxide beads in ACN and introducing them into empty spin columns. The reactors were then centrifuged to remove the solvent and washed thoroughly with DMF. Prior to their introduction to the reactors, the samples were resuspended in a solution composed of 5 μ L of water, 70 μ L of DMF and 25 μ L of methyl iodide. After allowing the samples to react for 15 min, the reactors were centrifuged at a slow speed, a second 25- μ L aliquot of methyl iodide was added, and the samples were re-applied to the reactors for a second 15-min reaction time. Next, the reactors were centrifuged and washed twice with 100- μ L aliquots of ACN. The permethylated glycans were then recovered from the reaction mixture by a liquid–liquid extraction with chloroform and a repeated washing with a 0.5-M NaCl solution and high-performance liquid chromatography-grade water. The samples were then dried in a vacuum centrifuge.

MALDI time-of-flight mass-spectrometric analysis

All of the samples were reconstituted in a 5- μ L aliquot of an 80%/20% water/methanol solution. The MALDI matrix solution, DHB, was prepared at a concentration of 10 mg/mL, supplemented with 1 mM sodium acetate, in an 80%/20% water/methanol solution. A 0.5- μ L aliquot of each sample was spotted in triplicate onto a standard MALDI plate and allowed to dry. Subsequently, a 0.4- μ L aliquot of the matrix solution was spotted and dried under vacuum, resulting in a thin layer of fine crystals. The spotted samples were interrogated with an Applied Biosystems 4800 MALDI time-of-flight (TOF)/TOF mass spectrometer (Forster City, CA). The instrument was operated in its positive-ion mode and a total of 1000 laser shots were acquired for each sample spot.

Data processing and statistical analysis

The acquired spectra were first base-line corrected and further processed by applying a noise filter using Data Explorer (version 4.0), a software tool which is included with the instrument's software package. Using the same tool, the spectra were exported as text files for further processing. The data were then normalized by expressing the relative intensity of each ion corresponding to a glycan as a percentage of the total intensity for all glycans included in this study. The average relative intensity was measured over the three spectra for each sample and this average was used for further data analysis. These normalized data were subjected to the battery of statistical tests. The different glycan classes, based on their structural characteristics, and later, the individual glycans were assessed for their diagnostic potential by first performing a one-way ANOVA test utilizing Microsoft Excel. The glycan classes and specific structures that resulted in a statistically significant *P*-value (<0.05) were then subjected to an ROC test using Origin 8.5 (OriginLabs Corp., Northampton, MA), resulting in an AUC value. In this statistical procedure, an AUC value of 1 indicates a perfectly positive test for the condition, whereas a score of zero indicates a perfectly negative result. Using a slight modification to the rubric as suggested previously (Swets 1988), the significance of the remaining AUC values were assigned. In this study, AUC values between 0.9 and 1 (positive for the disease) or 0 and 0.1 were considered highly accurate. Accurate tests were those that resulted in AUC values between 0.8 and 0.9 or 0.1 and 0.2. If the AUC value was between 0.7 and 0.8 or 0.2 and 0.3, the test was deemed as moderately accurate. Uninformative tests were those that produced an AUC value between 0.3 and 0.5 or from 0.5 to 0.7.

Supplementary data

Supplementary data for this article is available online at <http://glycob.oxfordjournals.org/>.

Funding

This work was supported by U01 CA128535-03 from the National Cancer Institute and GM24349 from the National Institute of General Medical Sciences, U.S. Department of Health and Human Services. The authors also thank the

Hoosier Oncology Group for providing the samples for this study and the support from the Walther Cancer Institute Foundation.

Conflict of interest

None declared.

Abbreviations

ACN, acetonitrile; AGP, α_1 -acid glycoprotein; ANOVA, analysis of variance; AUC, area under the curve; CT, computed tomography; DHB, 2,5-dihydroxybenzoic acid; DMF, dimethylformamide; DMT-MM, 4-(4,6-dimethoxy-1,3,5-triazin-2yl)-4-methylmorpholinium chloride; GlcNAc, *N*-acetylglucosamine; IL, interleukin; MALDI TOF, matrix-assisted laser desorption/ionization time-of-flight; MS, mass spectrometry; NSCLC, non-small cell lung carcinoma; PCA, principal component analysis; PNGase F, peptide-*N*-glycosidase F; ROC, receiver operator characteristics; SDS, sodium dodecyl sulfate; TFAA, trifluoroacetic acid.

References

- Aberle DR, Adams AM, Berge CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* 365:395–409.
- Alleva DG, Burger CJ, Elgert KD. 1994. Tumor-induced regulation of suppressor macrophage nitric oxide and TNF-alpha production. Role of tumor-derived IL-10, TGF-beta, and prostaglandin E2. *J Immunol.* 153:1674–1686.
- Alley WR, Jr, Madera M, Mechref Y, Novotny MV. 2010. Chip-based reversed-phase liquid chromatography-mass spectrometry of permethylated N-linked glycans: A potential methodology for cancer-biomarker discovery. *Anal Chem.* 82:5095–5106.
- Alley WR, Jr, Mann BF, Novotny MV. 2012. Analytical glycobiology at high sensitivity: Current approaches and directions. *Glycoconj J.* Submitted.
- Alley WR, Jr, Novotny MV. 2010. Glycomic analysis of sialic acid linkages in glycans derived from blood serum glycoproteins. *J Proteome Res.* 9:3062–3072.
- Alley WR, Jr, Vasseur JA, Goetz JA, Svoboda M, Mann BF, Matei DE, Menning N, Hussein A, Mechref Y, Novotny MV. 2012. N-linked glycan structures and their expressions change in the blood sera of ovarian cancer patients. *J Proteome Res.* 11:2282–2300.
- Andersen JD, Boylan KL, Xue FS, Anderson LB, Witthuhn BA, Markowski TW, Higgins L, Skubitz AP. 2010. Identification of candidate biomarkers in ovarian cancer serum by depletion of highly abundant proteins and differential in-gel electrophoresis. *Electrophoresis.* 31:599–610.
- Anderson NL, Anderson NG. 2002. The human plasma proteome. *Mol Cell Proteomics.* 1:845–867.
- Arnold JN, Saldova R, Galligan MC, Murphy TB, Mimura-Kimura Y, Telford JE, Godwin AK, Rudd PM. 2011. Novel glycan biomarkers for the detection of lung cancer. *J Proteome Res.* 10:1755–1764.
- Arnold JN, Saldova R, Hamid UM, Rudd PM. 2008. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics.* 8:3284–3293.
- Azuma Y, Murata M, Matsumoto K. 2000. Alteration of sugar chains on alpha(1)-acid glycoprotein secreted following cytokine stimulation of HuH-7 cells in vitro. *Clin Chim Acta.* 294:93–103.
- Bones J, Byrne JC, O'Donoghue N, McManus C, Scaife C, Boissin H, Nastase A, Rudd PM. 2011. Glycomic and glycoproteomic analysis of serum from patients with stomach cancer reveals potential markers arising from host defense response mechanisms. *J Proteome Res.* 10:1246–1265.
- Boyle P, Levin BE. 2008. World Cancer Report 2008. World Health Organization.

- Bradley WP, Blasco AP, Weiss JF, Alexander JC, Jr, Silverman NA, Chretien PB. 1977. Correlations among serum protein-bound carbohydrates, serum glycoproteins, lymphocyte reactivity, and tumors burden in cancer patients. *Cancer*. 40:2264–2272.
- Bruno R, Olivares R, Berille J, Chaikin P, Vivier N, Hammershaimb L, Rhodes GR, Rigas JR. 2003. Alpha-1-acid glycoprotein as an independent predictor for treatment effects and a prognostic factor of survival in patients with non-small cell lung cancer treated with docetaxel. *Clin Cancer Res*. 9:1077–1082.
- Davidoff GN, Votaw ML, Coon WW, Hultquist DE, Filter BJ, Wexler SA. 1978. Elevations in serum copper, erythrocytic copper, and ceruloplasmin concentrations in smokers. *Am J Clin Pathol*. 70:790–792.
- Delmotte P, Degroote S, Lafitte JJ, Lamblin G, Perini JM, Roussel P. 2002. Tumor necrosis factor alpha increases the expression of glycosyltransferases and sulfotransferases responsible for the biosynthesis of sialylated and/or sulfated Lewis x epitopes in the human bronchial mucosa. *J Biol Chem*. 277:424–431.
- Enewold L, Mechanic LE, Bowman ED, Zheng YL, Yu Z, Trivers G, Alberg AJ, Harris CC. 2009. Serum concentrations of cytokines and lung cancer survival in African Americans and Caucasians. *Cancer Epidemiol Biomarkers Prev*. 18:215–222.
- Goldman R, Resson HW, Varghese RS, Goldman L, Bascug G, Loffredo CA, Abdel-Hamid M, Gouda I, Ezzat S, Kyselova Z et al. 2009. Detection of hepatocellular carcinoma using glycomic analysis. *Clin Cancer Res*. 15:1808–1813.
- Gornik O, Lauc G. 2008. Glycosylation of serum proteins in inflammatory diseases. *Dis Markers*. 25:267–278.
- Gornik I, Maravic G, Dumic J, Fogel M, Lauc G. 1999. Fucosylation of IgG heavy chains is increased in rheumatoid arthritis. *Clin Biochem*. 32:605–608.
- Gornik O, Royle L, Harvey DJ, Radcliffe CM, Saldova R, Dwek RA, Rudd P, Lauc G. 2007. Changes of serum glycans during sepsis and acute pancreatitis. *Glycobiology*. 17:1321–1332.
- Hashimoto S, Asao T, Takahashi J, Yagihashi Y, Nishimura T, Saniabadi AR, Poland DC, van Dijk W, Kuwano H, Kochibe N et al. 2004. α 1-acid glycoprotein fucosylation as a marker of carcinoma progression and prognosis. *Cancer*. 101:2825–2836.
- Hauber HP, Foley SC, Hamid Q. 2006. Mucin overproduction in chronic inflammatory lung disease. *Can Respir J*. 13:327–335.
- Heo SH, Lee SJ, Ryoo HM, Park JY, Cho JY. 2007. Identification of putative serum glycoprotein biomarkers for human lung adenocarcinoma by multilectin affinity chromatography and LC-MS/MS. *Proteomics*. 7:4292–4302.
- Hoagland LF, Campa MJ, Gottlin EB, Herndon JE, 2nd, Patz EF, Jr. 2007. Haptoglobin and posttranslational glycan-modified derivatives as serum biomarkers for the diagnosis of nonsmall cell lung cancer. *Cancer*. 110:2260–2268.
- Huang Y, Konse T, Mechref Y, Novotny MV. 2002. Matrix-assisted laser desorption/ionization mass spectrometry compatible beta-elimination of O-linked oligosaccharides. *Rapid Commun Mass Spectrom*. 16:1199–1204.
- Imre T, Kremmer T, Heberger K, Molnar-Szollosi E, Ludanyi K, Pocsfalvi G, Malorni A, Drahos L, Vekey K. 2008. Mass spectrometric and linear discriminant analysis of N-glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals. *J Proteomics*. 71:186–197.
- Kang P, Madera M, Alley WR, Jr, Goldman R, Mechref Y, Novotny MV. 2011. Glycomic alterations in the highly-abundant and lesser-abundant blood serum protein fractions for patients diagnosed with hepatocellular carcinoma. *Int J Mass Spectrom*. 305:185–198.
- Kang P, Mechref Y, Klouckova I, Novotny MV. 2005. Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun Mass Spectrom*. 19:3421–3428.
- Kassis ES, Vaporciyan AA, Swisher SG, Correa AM, Bekele BN, Erasmus JJ, Hofstetter WL, Komaki R, Mehran RJ, Moran CA et al. 2009. Application of the revised lung cancer staging system (IASLC Staging Project) to a cancer center population. *J Thorac Cardiovasc Surg*. 138:412–418.e1–2.
- Knezevic A, Gornik O, Polasek O, Pucic M, Redzic I, Novokmet M, Rudd PM, Wright AF, Campbell H, Rudan I et al. 2010. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology*. 20:959–969.
- Knezevic A, Polasek O, Gornik O, Rudan I, Campbell H, Hayward C, Wright A, Kolcic I, O'Donoghue N, Bones J et al. 2009. Variability, heritability and environmental determinants of human plasma N-glycome. *J Proteome Res*. 8:694–701.
- Kossowska B, Ferens-Sieczkowska M, Gancarz R, Passowicz-Muszynska E, Jankowska R. 2005. Fucosylation of serum glycoproteins in lung cancer patients. *Clin Chem Lab Med*. 43:361–369.
- Kunishima M, Kawachi C, Iwasaki F, Terao K, Tani S. 1999. Synthesis and characterization of 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride. *Tetrahedron Lett*. 40:5327–5330.
- Kyselova Z, Mechref Y, Al Bataineh MM, Dobrolecki LE, Hickey RJ, Vinson J, Sweeney CJ, Novotny MV. 2007. Alterations in the serum glycome due to metastatic prostate cancer. *J Proteome Res*. 6:1822–1832.
- Kyselova Z, Mechref Y, Kang P, Goetz JA, Dobrolecki LE, Sledge GW, Schnaper L, Hickey RJ, Malkas LH, Novotny MV. 2008. Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. *Clin Chem*. 54:1166–1175.
- Lacunza I, Kremmer T, Diez-Masa JC, Sanz J, de Frutos M. 2007. Comparison of alpha-1-acid glycoprotein isoforms from healthy and cancer patients by capillary IEF. *Electrophoresis*. 28:4447–4451.
- Lattova E, Tomanek B, Bartusik D, Perreault H. 2010. N-glycomic changes in human breast carcinoma MCF-7 and T-lymphoblastoid cells after treatment with herceptin and herceptin/Lipoplex. *J Proteome Res*. 9:1533–1540.
- Lin Z, Simeone DM, Anderson MA, Brand RE, Xie X, Shedden KA, Ruffin MT, Lubman DM. 2011. Mass spectrometric assay for analysis of haptoglobin fucosylation in pancreatic cancer. *J Proteome Res*. 10:2602–2611.
- Lin B, White JT, Wu J, Lele S, Old LJ, Hood L, Odunsi K. 2009. Deep depletion of abundant serum proteins reveals low-abundant proteins as potential biomarkers for human ovarian cancer. *Proteomics Clin Appl*. 3:853–861.
- Malhotra R, Wormald MR, Rudd PM, Fischer PB, Dwek RA, Sim RB. 1995. Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. *Nat Med*. 1:237–243.
- Mechref Y, Hussein A, Bekesova S, Pungpapong V, Zhang M, Dobrolecki LE, Hickey RJ, Hammoud ZT, Novotny MV. 2009. Quantitative serum glycomics of esophageal adenocarcinoma and other esophageal disease onsets. *J Proteome Res*. 8:2656–2666.
- Mehta A, Block TM. 2008. Fucosylated glycoproteins as markers of liver disease. *Dis Markers*. 25:259–265.
- Moszczynski P, Zabinski Z, Moszczynski P, Jr, Rutowski J, Slowinski S, Tabarowski Z. 2001. Immunological findings in cigarette smokers. *Toxicol Lett*. 118:121–127.
- Mulshine JL, Sullivan DC. 2005. Clinical practice. Lung cancer screening. *N Engl J Med*. 352:2714–2720.
- Nakao H, Nishikawa A, Karasuno T, Nishiura T, Iida M, Kanayama Y, Yonezawa T, Tarui S, Taniguchi N. 1990. Modulation of N-acetylglucosaminyltransferase III, IV and V activities and alteration of the surface oligosaccharide structure of a myeloma cell line by interleukin 6. *Biochem Biophys Res Commun*. 172:1260–1266.
- Ogawa J, Inoue H, Koide S. 1996. Expression of alpha-1,3-fucosyltransferase type IV and VII genes is related to poor prognosis in lung cancer. *Cancer Res*. 56:325–329.
- Okada M, Nishio W, Sakamoto T, Uchino K, Yuki T, Nakagawa A, Tsubota N. 2005. Effect of tumor size on prognosis in patients with non-small cell lung cancer: The role of segmentectomy as a type of lesser resection. *J Thorac Cardiovasc Surg*. 129:87–93.
- Olewicz-Gawlik A, Korczowska-Lacka I, Lacki JK, Klama K, Hrycaj P. 2007. Fucosylation of serum alpha-1-acid glycoprotein in rheumatoid arthritis patients treated with infliximab. *Clin Rheumatol*. 26:1679–1684.
- Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, Miller YE, Pass HI, Rom WN, Siegfried JM, Stewart A et al. 2010. Unlocking biomarker discovery: Large scale application of aptamer proteomic technology for early detection of lung cancer. *PLoS One*. 5:e15003.
- Rudd P, Fortune F, Lehner T, Parekh R, Patel T, Wormald M, Malhotra R, Sim R, Dwek R. 1995. Lectin-carbohydrate interactions in disease. T-cell recognition of IgA and IgD; mannose binding protein recognition of IgG0. *Adv Exp Med Biol*. 376:147–152.
- Saldova R, Royle L, Radcliffe CM, Abd Hamid UM, Evans R, Arnold JN, Banks RE, Hutson R, Harvey DJ, Antrobus R et al. 2007. Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiology*. 17:1344–1356.
- Saldova R, Wormald MR, Dwek RA, Rudd PM. 2008. Glycosylation changes on serum glycoproteins in ovarian cancer may contribute to disease pathogenesis. *Dis Markers*. 25:219–232.
- Schulz BL, Sloane AJ, Robinson LJ, Sebastian LT, Glanville AR, Song Y, Verkman AS, Harry JL, Packer NH, Karlsson NG. 2005. Mucin glycosylation changes in cystic fibrosis lung disease are not manifest in submucosal gland secretions. *Biochem J*. 387:911–919.

- Shima M, Adachi M. 1996. Effects of environmental tobacco smoke on serum levels of acute phase proteins in schoolchildren. *Prev Med.* 25:617–624.
- Shiyan SD, Bovin NV. 1997. Carbohydrate composition and immunomodulatory activity of different glycoforms of alpha1-acid glycoprotein. *Glycoconj J.* 14:631–638.
- Smith MP, Wood SL, Zougman A, Ho JT, Peng J, Jackson D, Cairns DA, Lewington AJ, Selby PJ, Banks RE. 2011. A systematic analysis of the effects of increasing degrees of serum immunodepletion in terms of depth of coverage and other key aspects in top-down and bottom-up proteomic analyses. *Proteomics.* 11:2222–2235.
- Svoboda M, Mann BF, Goetz JA, Novotny MV. 2012. Examination of glycan profiles from IgG-depleted human immunoglobulins facilitated by micro-scale affinity chromatography. *Anal Chem.* 84:3269–3277.
- Swets JA. 1988. Measuring the accuracy of diagnostic systems. *Science.* 240:1285–1293.
- Takizawa H, Ohtoshi T, Ohta K, Yamashita N, Hirohata S, Hirai K, Hiramatsu K, Ito K. 1993. Growth inhibition of human lung cancer cell lines by interleukin 6 in vitro: A possible role in tumor growth via an autocrine mechanism. *Cancer Res.* 53:4175–4181.
- Tang Z, Varghese RS, Bekesova S, Loffredo CA, Abdul-Hamid M, Kyselova Z, Mechref Y, Novotny MV, Goldman R, Ransom HW. 2010. Identification of N-glycan serum markers associated with hepatocellular carcinoma from mass spectrometry data. *J Proteome Res.* 9:104–112.
- Tsai HY, Boonyapranai K, Sriyam S, Yu CJ, Wu SW, Khoo KH, Phutrakul S, Chen ST. 2011. Glycoproteomics analysis to identify a glycoform on haptoglobin associated with lung cancer. *Proteomics.* 11:2162–2170.
- Ueda K, Fukase Y, Katagiri T, Ishikawa N, Irie S, Sato TA, Ito H, Nakayama H, Miyagi Y, Tsuchiya E et al. 2009. Targeted serum glycoproteomics for the discovery of lung cancer-associated glycosylation disorders using lectin-coupled ProteinChip arrays. *Proteomics.* 9:2182–2192.
- Van Den Heuvel MM, Poland DC, De Graaff CS, Hoefsmit EC, Postmus PE, Beelen RH, Van Dijk W. 2000. The degree of branching of the glycans of alpha(1)-acid glycoprotein in asthma. A correlation with lung function and inflammatory parameters. *Am J Respir Crit Care Med.* 161:1972–1978.
- Voynow JA, Gendler SJ, Rose MC. 2006. Regulation of mucin genes in chronic inflammatory airway diseases. *Am J Respir Cell Mol Biol.* 34:661–665.
- Weiss JF, Wolf GT, Edwards BK, Chretien PB. 1981. Effects of smoking and age on serum levels of immune-reactive proteins altered in cancer patients. *Cancer Detect Prev.* 4:211–217.
- Welch HG, Woloshin S, Schwartz LM, Gordis L, Gotzche PC, Harris R, Kramer BS, Ransohoff DF. 2007. Overstating the evidence for lung cancer screening: The International Early Lung Cancer Action Program (I-ELCAP) Study. *Arch Intern Med.* 167:2289–2295.
- Wilson DO, Weissfeld JL, Furlman CR, Fisher SN, Balogh P, Landreneau RJ, Luketich JD, Siegfried JM. 2008. The Pittsburgh Lung Screening Study (PLUSS): Outcomes within 3 years of a first computed tomography scan. *Am J Respir Crit Care Med.* 178:956–961.
- Yao M, Zhou DP, Jiang SM, Wang QH, Zhou XD, Tang ZY, Gu JX. 1998. Elevated activity of N-acetylglucosaminyltransferase V in human hepatocellular carcinoma. *J Cancer Res Clin Oncol.* 124:27–30.
- Zeng Z, Hincapie M, Pitteri SJ, Hanash S, Schalkwijk J, Hogan JM, Wang H, Hancock WS. 2011. A proteomics platform combining depletion, multi-lectin affinity chromatography (M-LAC), and isoelectric focusing to study the breast cancer proteome. *Anal Chem.* 83:4845–4854.
- Zeng X, Hood BL, Sun M, Conrads TP, Day RS, Weissfeld JL, Siegfried JM, Bigbee WL. 2010. Lung cancer serum biomarker discovery using glycoprotein capture and liquid chromatography mass spectrometry. *J Proteome Res.* 9:6440–6449.