Diagnostic and Prognostic Performance of Blood Plasma Glycan Features in the Women Epidemiology Lung Cancer (WELCA) study

Yueming Hu^a, Shadi Ferdosi^{a, †}, Erandi P. Kapuruge^a, Jesús Aguilar Diaz de Leon^a, Isabelle Stücker^b, Loredana Radob^{b,c}, Pascal Guénel^b, and Chad R. Borges^{a,*}

^a School of Molecular Sciences and The Biodesign Institute at Arizona State University, Tempe, AZ, 85281

^b CESP (Center for Research in Epidemiology and Population Health), Cancer and Environment team, INSERM UMS1018, University Paris-Sud, University Paris-Saclay, Villejuif, France

^c University Paris Descartes, Faculty of Dental Surgery, Paris, France

* Author to whom correspondence should be addressed: Chad R. Borges, The Biodesign Institute at Arizona State University, P.O. Box 876401, Tempe, AZ 85287. Tel 480-727-9928; email:

chad.borges@asu.edu

[†]Current address: Seer INC, South San Francisco, CA, 94080

Table of Contents

Table S1 (Page S-4): Basic Clinical Characteristics and n-values of the WELCA Sample Set.

Table S2 (Page S-5): Stage and Gender Composition of Three Lung Cancer Sample Sets and Their Sub-Cohorts.

Table S3 (Page S-6): Comparison of Glycan Node Stability at Different Conditions Relative to Control Aliquots Stored at -80 °C.

Table S4 (Page S-7): Statistically Significant Differences between Cohorts within the WELCA Study.

Table S5 (Page S-8 to S-9): Stage-by-Stage ROC Comparison of the Top Performing Glycan Nodes.

Table S6 (Page S-10 to S-11): Comparison of Top Performing Glycan Nodes in Male vs. Female Patients with Early Stage Lung Cancer.

Table S7 (Page S-12): Correlation Between Age and the Top Performing Glycan Nodes in the WELCA Cases (all stages) and, Separately, Controls.

Table S8 (Page S-13 to S-14): Comparison of the Top Performing Glycan Nodes in Different Histological Types.

Table S9 (Page S-15 to S-16): Stage-by-Stage Comparison of Total Glycosylation withIndividual Glycan Feature.

Table S10 (Page S-17 to S-18): Survival Prediction by the Top Performing Glycan Nodes in All Stages, Stage III and IV Combined, Stage III Only and Stage IV Only.

Figure S1 (Page S-19) ROC curves for β 1-4 branching for stage IV vs. each other stage of nonsmall cell lung cancer (NSCLC).

Figure S2 (Page S-20): Connection between antennary fucosylation and smoking status within the WELCA control group.

Figure S3 (Page S-21): ROC curves for the six top performing glycan nodes within different histological subtypes of non-small cell lung cancer (NSCLC).

Figure S4 (Page S-22): Multivariate logistic regression models for stage I–IV patients from the WELCA data set.

Figure S5 (Page S-23): Survival curves of the six top performing glycan nodes for stage III and IV combined.

Figure S6 (Page S-24 to S-25): Survival curves for the two top performing glycan nodes that were significantly different between stages III and IV: Stage III patients alone and stage IV patients alone.

Hu et al_Raw Data.xlsm – Raw and normalized chromatographic peak areas for all WELCA samples analyzed in this study.

		W	ELCA set
		Controls	Cases
Age ^a		61.2 ± 9.73	61.6 ± 9.04
Smoking Status	Never-Smoker	90	36
	Previous-Smoker	72	52
	Current-Smoker	45	98
	Unknown	0	22
Staging	Stage I	N/A	16
	Stage II	N/A	13
	Stage III	N/A	45
	Stage IV	N/A	99
	Unknown Stage	N/A	35
Tumor Histological	SCLC ^b – Located	N/A	7
i ypes	SCLC - Disseminated	N/A	12
	NSCLC ^c - Adenocarcinoma	N/A	131
	NSCLC - Squamous cell carcinoma	N/A	25
	NSCLC - Large cell carcinoma	N/A	13
	NSCLC - Adenosquamous	N/A	1
	NSCLC - Sarcoma	N/A	3
	Unknown	N/A	16

Table S1. Basic Clinical Characteristics and n-values of the WELCA Sample Set.

^{*a*}Age in years \pm SD.

^b Small cell lung cancer

^c non-small cell lung cancer

Table S2.	Stage and	Gender	Composition	of Three	Lung	Cancer	Sample Set	s and	Their
Sub-Coho	orts.								

Name of Sample Set	Plasma or Serum	Controls (M/F)	Stage I (M/F)	Stage II (M/F)	Stage III (M/F)	Stage IV (M/F)
WELCA set	plasma	0/207	0/16	0/13	0/45	0/99
Dual Gender Lung Cancer set	plasma	123/76	14/6	12/8	50/31	47/31
Stage I Only Lung Cancer set	serum	28/45	33/74	-	-	-

Table S3.	. Comparison	of Glycan	Node Stabilit	ty at Different	t Conditions	Relative to	Control
Aliquots	Stored at -80 °	PC. ^{<i>a</i>}					

Glycan	10 days at	90 days at	360 days	2 days at	90 days at	1 day at
Node	-20 °C	-20 °C	at -20 °C	4 °C	4 °C	25 °C
t-Fucose	ns	ns	ns	ns	ns	ns
t-Gal	ns	ns	ns	ns	ns	ns
2-Man	ns	ns	ns	ns	ns	ns
4-Glc	ns	ns	ns	ns	ns	ns
3-Gal	ns	ns	ns	ns	ns	ns
6-Gal	ns	ns	ns	ns	ns	*
2,4-Man	ns	ns	ns	ns	ns	ns
2,6-Man	ns	ns	ns	ns	ns	ns
3,6-Man	ns	ns	ns	ns	ns	ns
3,4,6-Man	ns	ns	ns	ns	ns	ns
t-GlcNAc	ns	ns	ns	ns	ns	ns
4-GlcNAc	ns	ns	ns	ns	ns	ns
3-GalNAc	ns	ns	ns	ns	ns	ns
3,4-GlcNAc	ns	ns	ns	ns	ns	ns
4,6-GlcNAc	ns	ns	ns	ns	ns	ns

^{*a*}Heavy, stable isotope labeled glucose (Glc) and GlcNAc were utilized to normalize Hexose and HexNAc data, correspondingly.

^{*b*}Results of Friedman test followed by Dunn's post hoc test at 95% confidence level are given. "ns" stands for "not significant". "*" indicates p < 0.05.

Glycan Node ^b	Control vs Stage I	Control vs Stage II	Control vs Stage III	Control vs Stage IV	Stage I vs Stage II	Stage I vs Stage III	Stage I vs Stage IV	Stage II vs Stage III	Stage II vs Stage IV	Stage III vs Stage IV
t-Fucose	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
t-Gal	ns	ns	dddd	dddd	ns	dd	ns	ns	ns	ns
2-Man	ns	ns	ns	dd	ns	ns	ns	ns	d	d
4-Glc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3-Gal	ns	dd	dd	ns	ns	ns	ns	ns	ns	ns
6-Gal	ns	ns	i	iiii	ns	ns	ii	ns	ns	ns
3,4-Gal	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
2,4-Man	ns	ns	ii	iiii	ns	ns	iiii	ns	ns	iiii
2,6-Man	ns	ii	iii	iiii	ns	ns	ns	ns	ns	ns
3,6-Man	ns	dd	ns	ns	dd	ns	ns	ns	ii	ns
3,6-Gal	ns	dd	ns	ns	ns	ns	ns	ns	ii	ns
3,4,6-Man	ns	dd	ddd	dddd	d	ns	ddd	ns	ns	dd
t-GlcNAc	ns	d	dd	dddd	ns	ns	ddd	ns	ns	ddd
4-GlcNAc	ns	ns	ns	iiii	ns	ns	ns	ns	ns	ns
3-GlcNAc	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3-GalNAc	ns	ns	dddd	dddd	ns	ns	dd	ns	ns	ns
3,4-GlcNAc	ns	ii	iiii	iiii	i	i	ii	ns	ns	ns
4,6-GlcNAc	ns	d	dd	dddd	dd	dd	dddd	ns	ns	d
3,6-GalNAc	ns	ns	dd	dddd	ns	ns	ns	ns	ns	ns

Table S4. Statistically Significant Differences between Cohorts within the WELCA Study^a.

^{*a*}Hexose data were normalized to the sum of endogenous hexoses, and HexNAc data were normalized to the sum of endogenous HexNAcs.

^{*b*}Kruskal-Wallis test followed by Benjamini–Hochberg false discovery correction procedure at 95% confidence level is given. "ns" stands for "not significant". "i" and "d" stands for "increased" and "decreased". i/d indicates p < 0.05. ii/dd indicates p < 0.01. iii/ddd indicates p < 0.001, and iiii/ddd indicates p < 0.0001.

Stages ^a	Glycan Feature	A: ROC AUC of set A	<i>p</i> -value of Delong's test for two ROC curves ^b
		B: ROC AUC of set B	
Stage I	$\alpha 2$ - 6 Sialylation	A: 0.733	0.031 (NS)
		B: 0.564	-
Set A: WELCA Set	β1 - 4 Branching	A: 0.696	0.112 (NS)
Set B: Stage I-Only Lung Cancer Set		B: 0.549	-
	β1 - 6 Branching	A: 0.797	0.008
		B: 0.592	-
	Antennary Fucosylation	A: 0.609	0.965 (NS)
		B: 0.613	-
Stage I	α2-6 Sialylation	A: 0.733	0.092 (NS)
		B: 0.575	-
Set A: WELCA Set	β1 - 4 Branching	A: 0.696	0.264 (NS)
Set B: Dual Gender Lung Cancer Set		B: 0.579	-
	β1 - 6 Branching	A: 0.797	0.008
		B: 0.547	-
	Antennary Fucosylation	A: 0.609	0.885 (NS)
	i deosylution	B: 0.594	-
Stage II	α2-6 Sialylation	A: 0.681	0.509 (NS)
		B: 0.607	
Set A: WELCA Set	β1 - 4 Branching	A: 0.707	0.489 (NS)
Set B: Dual Gender Lung Cancer Set		B: 0.630	
	β1 - 6 Branching	A: 0.770	0.071 (NS)

Table S5. Stage-by-Stage ROC Comparison of the Top Performing Glycan Nodes.

		B: 0.582	
	Antennary Fucosylation	A: 0.760	0.302 (NS)
	1 deosylation	B: 0.655	
Stage III	α2-6 Sialylation	A: 0.796	0.241 (NS)
		B: 0.739	
Set A: WELCA Set	β1 - 4 Branching	A: 0.798	0.407 (NS)
Set B: Dual Gender		B: 0.755	
Lung Cuncer Set	β1 - 6 Branching	A: 0.822	0.119 (NS)
		B: 0.745	
	Antennary	A: 0.826	0.161 (NS)
	1 deosylation	B: 0.754	
Stage IV	α2-6 Sialylation	A: 0.887	0.009
		B: 0.791	
Set A: WELCA Set	β1 - 4 Branching	A: 0.917	0.002
Set B: Dual Gender Lung Cancer Set		B: 0.802	
	β1 - 6 Branching	A: 0.907	0.008
		B: 0.810	
	Antennary Fucosylation	A: 0.822	0.307 (NS)
	Fucosylation	B: 0.777	

^{*a*}The WELCA set was compared to the Dual Gender Lung Cancer Set and Stage I-Only Lung Cancer Set (also dual gender). *N*-values of each group are shown in Table S2. Actual ROC curves are shown in Fig. 4.

^b"NS" indicates no significant difference between the two compared ROC curves. The significant levels of p values are adjusted by Bonferroni multiple comparison correction: p > 0.013 (NS), p < 0.013 (*), p < 0.003 (**), p < 0.003 (***).

Table S6. Comparison of Top Performing Glycan Nodes in Male vs. Female Patients withEarly Stage Lung Cancer.

Lung Cancer Sets ^a	Glycan Feature	A: ROC AUC of set	<i>p</i> -value of Delong's test for two ROC curves ^b
		B: KOC AUC of set B	
Dual Gender Lung Cancer Set	2-linked Mannose	A: 0.618	0.537 (NS)
Stage I		B: 0.544	
Male vs Female	α 2 - 6 Sialylation	A: 0.646	0.393 (NS) ^c
		B: 0.553	
Set A:	β1 - 4 Branching	A: 0.629	0.464 (NS) ^c
Male Patients $(n = 14)$		B: 0.539	
vs Male Controls (n = 123) Set B:	β1 - 6 Branching	A: 0.598	0.870 (NS) ^c
Female Patients (n = 6)		B: 0.579	
vs Female Controls (n = 76)	4-linked GlcNAc	A: 0.592	0.686 (NS)
		B: 0.537	
		A: 0.552	0.206 (NS)
	1 deosylation	B: 0.702	
Dual Gender Lung Cancer Set	2-linked Mannose	A: 0.628	0.719 (NS)
Stage II		B: 0.579	
Male vs Female	α 2 - 6 Sialylation	A: 0.626	0.873 (NS)
		B: 0.602	
Set A:	β1 - 4 Branching	A: 0.633	0.918 (NS)
Male Patients (n = 12)		B: 0.618	
vs Male Controls (n = 123) Set B.	β1 - 6 Branching	A: 0.594	0.826 (NS)
501 D .		B: 0.559	

Female Patients (n = 8)	4-linked GlcNAc	A: 0.648	0.533 (NS)
vs Female Controls (n = 76)		B: 0.564	
	Antennary Fucosylation	A: 0.737	0.382 (NS)
	i deosylution	B: 0.615	
Stage I-Only Lung Cancer Set	2-linked Mannose	A: 0.655	0.247 (NS)
Male vs Female		B: 0.547	
	α2-6 Sialylation	A: 0.585	0.766 (NS)
Set A:		B: 0.557	
Male Patients (n = 33)	β1 - 4 Branching	A: 0.448	0.219 (NS) ^c
vs Male Controls (n = 28)		B: 0.563	
Female Patients (n = 74)	β1 - 6 Branching	A: 0.450	0.0825 (NS) ^c
vs Female Controls (n = 45)		B: 0.616	
	4-linked GlcNAc	A: 0.650	0.508 (NS)
		B: 0.589	
	Antennary Fucosylation	A: 0.632	0.811 (NS)
		B: 0.610	

^{*a*}Comparisons are made for stage I and II of the Dual Gender Lung Cancer Set, and the Stage I Only Lung Cancer Set. Unpaired Delong's test or Bootstrap test are applied to compare two ROC curves.

^{*b*"NS" indicates no significant difference between the two compared ROC curves. The significant levels of *p* values are adjusted by Bonferroni multiple comparison correction: p > 0.0083 (NS), p < 0.0083 (*), p < 0.0017 (**), p < 0.00017 (***).}

^c*p*-value is from Bootstrap test instead of Delong's test, because Delong's test should not be applied to ROC curves with different directions and the stratification of Bootstrap is especially useful if groups are not balanced.

Table S7. Correlation Between Age and the Top Performing Glycan Nodes in the WELCACases (all stages) and, Separately, Controls.

Case/Control ^a	Glycan Feature	Correlation coefficient (r)	<i>p</i> -value of Spearman's rank correlation ^b
Case	2-linked Mannose	0.102	0.168 (NS)
	$\alpha 2$ - 6 Sialylation	0.073	0.324 (NS)
n = 208	β1 - 4 Branching	0.072	0.329(NS)
	β1 - 6 Branching	0.091	0.217 (NS)
	4-linked GlcNAc	0.030	0.681 (NS)
	Antennary Fucosylation	0.148	0.044 (NS)
Control	2-linked Mannose	0.039	0.577 (NS)
	α2-6 Sialylation	0.075	0.283 (NS)
n = 207	β1 - 4 Branching	0.047	0.501 (NS)
	β1 - 6 Branching	0.103	0.142(NS)
	4-linked GlcNAc	-0.101	0.148 (NS)
	Antennary Fucosylation	0.205	0.0031

^{*a*}Spearman's rank correlation coefficients and *p* values are provided for the six top performing glycan features in all cases (n = 208) and controls (n = 207).

^{*b*"NS" indicates no significant correlation between age and the corresponding glycan feature. The significant levels of *p* values are adjusted by Bonferroni multiple comparison correction: p > 0.0083 (NS), p < 0.0083 (*), p < 0.0017(**).}

Histological Types ^a	Glycan Feature	A: ROC AUC of set	<i>p</i> -value of Delong's test for two ROC curves ^b
		B: KOC AUC of set B	
Adenocarcinoma	2-linked Mannose	A: 0.854	0.071 (NS) ^c
vs		B: 0.926	
Squamous cell carcinoma	α 2 - 6 Sialylation	A: 0.878	0.130 (NS) ^c
curemoniu		B: 0.939	
Set A:	β1 - 4 Branching	A: 0.908	0.114 (NS) ^c
Adenocarcinoma vs Controls		B: 0.960	
Set B. Squamous	β 1 - 6 Branching	A: 0.906	0.539 (NS) ^c
cell carcinoma vs Controls		B: 0.939	
	4-linked GlcNAc Antennary Fucosylation	A: 0.877	0.702 (NS) ^c
		B: 0.899	
		A: 0.815	0.608 (NS) ^c
		B: 0.861	
Adenocarcinoma	2-linked Mannose	A: 0.854	0.957 (NS) ^c
VS		B: 0.860	
Large cell carcinoma	α 2 - 6 Sialylation	A: 0.878	0.647 (NS) ^c
		B: 0.817	
Set A:	β 1 - 4 Branching	A: 0.908	0.586 (NS) ^c
Adenocarcinoma vs Controls		B: 0.828	
Set B: Large cell	β 1 - 6 Branching	A: 0.906	0.402 (NS) ^c
carcinoma vs		B: 0.808	

Table S8. Comparison of the Top Performing Glycan Nodes in Different HistologicalTypes.

Controls	4-linked GlcNAc	A: 0.877	0.934 (NS) ^c	
		B: 0.869		
	Antennary Fucosylation	A: 0.815	0.706 (NS) ^c	
		B: 0.757	•	
Squamous cell carcinoma	2-linked Mannose	A: 0.926	0.556 (NS)	
VS		B: 0.860		
Large cell carcinoma	α2-6 Sialylation	A: 0.939	0.406 (NS)	
		B: 0.817		
	β1 - 4 Branching	A: 0.960	0.426 (NS)	
Set A: Squamous cell carcinoma vs		B: 0.828		
Controls	β1 - 6 Branching	A: 0.939	0.332 (NS)	
Set B: Large cell		B: 0.808		
carcinoma vs Controls	4-linked GlcNAc	A: 0.899	0.804 (NS)	
		B: 0.869		
	Antennary Fucosylation	A: 0.861	0.578 (NS)	
	2 00005 100000	B: 0.757		

^{*a*}Comparisons are made for stage IV patients with various histological types of non-small cell lung cancer (NSCLC) vs. all controls. The n-values for the different histological sets are as following. Adenocarcinoma set: n = 70; Squamous cell carcinoma set: n = 8; Large cell carcinoma set: n = 5; Controls: n = 207. Unpaired Delong's test or Bootstrap test are used to compare two ROC curves.

^b"NS" indicates no significant difference between the two compared ROC curves. The significant levels of p values are adjusted by Bonferroni multiple comparison correction: p > 0.0083 (NS).

^c*p*-value is from Bootstrap test instead of Delong's test, because the stratification of Bootstrap is especially useful if groups are not balanced.

Stages ^a	Gly	ycan Feature	ROC AUC	p-value of Delong's test for two ROC curves ^b
Stage I	А	β1-6 Branching	0.797	A vs B: 0.280 (NS)
	В	Total Hexoses	0.750	A vs C: 0.196 (NS)
	С	Total HexNAcs	0.730	A ve D: 0.289 (NS)
	D	Total Hexoses and HexNAcs	0.755	A VS D. 0.269 (INS)
Stage II	A	β1-6 Branching	0.770	A vs B: 0.024 (NS)
	В	Total Hexoses	0.674	A vs C: 0.091 (NS)
	С	Total HexNAcs	0.627	A va D: 0.017
	D	Total Hexoses and HexNAcs	0.679	A VS D. 0.017
Stage III	А	2-linked Mannose	0.843	A vs B: 0.938 (NS)
	В	Total Hexoses	0.844	A vs C: 0.676 (NS)
	С	Total HexNAcs	0.830	
	D	Total Hexoses and HexNAcs	0.860	A vs D: 0.196 (NS)
Stage IV	А	β1-4 Branching	0.917	A vs B: 0.021 (NS)
	В	Total Hexoses	0.892	A vs C: 0.159 (NS)
	С	Total HexNAcs	0.891	

Table S9. Stage-by-Stage Comparison of Total Glycosylation with Individual GlycanFeature.

D	Total Hexoses	0.907	A vs D: 0.280 (NS)
	and HexNAcs		

^{*a*}For each stage, the individual top performing glycan node with the largest area under curve (AUC) value was selected to compare to total hexoses (sum of all hexose glycan nodes), total HexNAcs (sum of all HexNAc glycan nodes) and total Hexoses and HexNAcs (sum of all glycan nodes). A paired Delong's test was utilized to compare two ROC curves.

^b"NS" indicates no significant difference between the two compared ROC curves. The significant levels of p values are adjusted by Bonferroni multiple comparison correction: p > 0.017 (NS), p < 0.017 (*), p < 0.0033 (**).

Table S10. Survival Prediction by the Top Performing Glycan Nodes in All Stages, StageIII and IV Combined, Stage III Only and Stage IV Only.

Stage Involved	Glycan Feature	Cox proportional hazards regression model ^a			
mvorveu		<i>p</i> -value ^b	Hazard Ratio	Lower bound at 95% CL	Upper bound at 95% CL
All stages	2-linked Mannose	0.0003	2.39	1.49	3.83
	α 2 - 6 Sialylation	0.0002	2.48	1.53	4.03
n = 197	β1 - 4 Branching	< 0.0001	2.70	1.66	4.41
	β1 - 6 Branching	0.0002	2.54	1.57	4.12
	4-linked GlcNAc	0.0066	1.99	1.21	3.26
	Antennary Fucosylation	< 0.0001	2.75	1.70	4.42
Stage III & IV n = 138	2-linked Mannose	0.0059	2.09	1.24	3.52
	α 2 - 6 Sialylation	0.0029	2.16	1.30	3.58
	β1 - 4 Branching	0.0014	2.29	1.38	3.82
	β1 - 6 Branching	0.0014	2.29	1.38	3.82
	4-linked GlcNAc	0.0148 (NS)	1.98	1.14	3.42
	Antennary Fucosylation	0.0011	2.45	1.43	4.18
Stage III	2-linked Mannose	0.3291 (NS)	1.69	0.59	4.81
	α 2 - 6 Sialylation	0.1551 (NS)	2.12	0.75	5.99
n = 44	β1 - 4 Branching	0.5653 (NS)	1.35	0.49	3.75
	β1 - 6 Branching	0.1685 (NS)	2.04	0.74	5.65
	4-linked GlcNAc	0.2910 (NS)	1.68	0.64	4.42
	Antennary Fucosylation	0.0769 (NS)	2.61	0.90	7.58

Stage IV	2-linked Mannose	0.0131 (NS)	2.19	1.18	4.05
	α2-6 Sialylation	0.0051	2.42	1.30	4.50
n = 94	β1 - 4 Branching	0.0011	2.82	1.51	5.26
	β1 - 6 Branching	0.0032	2.53	1.36	4.67
	4-linked GlcNAc	0.0220 (NS)	2.19	1.12	4.31
	Antennary Fucosylation	0.0081	2.31	1.24	4.31

^{*a*}Cox proportional hazards regression model p values and hazard ratios for the top quartile for each glycan node vs. all other quartiles combined, and lower and upper bound at 95% confident limits of hazard ratios are provided.

^b"NS" indicates no statistically significance between hazard ratio and 1, representing no difference in the relative risk of death, comparing patients in the top quartile vs. all other quartiles of the respective glycan node. The significant levels of *p* values are adjusted by Bonferroni multiple comparison correction: p > 0.0083 (NS), p < 0.0083 (*), p < 0.0017 (**), p < 0.00017 (***).



Figure S1. ROC curves for β 1-4 branching for stage IV vs. each other stage of non-small cell lung cancer (NSCLC). *N*-values of each group are provided in **Table S1**. ROC curves for stage I-III lung cancer cases vs. controls are provided in panels a-c. Areas under the ROC curves (AUC) values and *p* values are provided under each ROC curve



Figure S2. Connection between antennary fucosylation and smoking status within the WELCA control group. (a) The univariate distributions of antennary fucosylation within the control group are shown, subdivided by smoking status. Different letters above the data points indicate statistically significant differences between groups as detected by the Kruskal-Wallis test followed by the Benjamini-Hochberg FDR correction procedure. Spearman's rank correlation between antennary fucosylation and smoking pack-years for (b) all control patients and (c) control patients with smoking history (smoking pack-year > 0). Correlation coefficients are provided above the data points. "NS" next to the correlation coefficient demonstrates a lack of statistically significant associations with smoking were found.



Figure S3. ROC curves for the six top performing glycan nodes within different histological subtypes of non-small cell lung cancer (NSCLC). *N*-values of each group are provided in **Table S8**. Results of unpaired Delong's test and Bootstrap test indicated no significant differences between ROC curves of different histology subtypes of NSCLC (see **Table S8**). ROC AUC values are provided in parenthesis next to the specified histological subtypes. "NS" next to the AUC values indicates no significant difference was found between cases and controls.



Figure S4. Multivariate logistic regression models for stage I–IV patients from the WELCA data set. Three multivariate logistic regression models were built and corresponding ROC curves were plotted for each stage with different fitting procedures: (1) fitted once on the complete data set and acquired probability (referred to as "Fitted Probabilities") with no use of cross-validation; (2) fitted once on the complete data set, cross-validated with fixed predictors but mobile parameter estimates at each iteration (predicted probability referred to as "CV Probabilities (semi)"); and (3) refitted at each iteration of cross-validation (corresponding probability demonstrated as "CV Probabilities (full)"). ROC AUC values are provided in parenthesis next to the specified models. For each stage, the ROC curve of the best performing individual glycan node was selected and compared to the fully validated multivariate model. No significant differences were detected (Delong's test).



Figure S5. Survival curves of the six top performing glycan nodes for stage III and IV combined. In each panel, the top quartile of specified glycan node is compared to all other quartiles combined. According to the results of a log-rank Mantel-Cox test, the survival curves within each panel are significantly different (p < 0.05). Dotted lines represent 95% confident intervals. The median duration of follow-up for patients that died, until death, was 393 days; for survivors this value was 1264 days. The median total follow-up time for all patients was 908 days.



Figure S6. Survival curves for the two top performing glycan nodes that were significantly different between stages III and IV: Stage III patients alone and stage IV patients alone. The top α 2-6 Sialylation quartile is compared to all other quartiles combined for stage III patients (panel a) and stage IV patients (panel c). Similarly, the top β 1-4 Branching quartile is compared to all other quartiles combined for stage III patients (b) and stage IV patients (d). In each plot, the *p* value of the log-rank Mantel-Cox test is provided, indicating whether significant differences were determined for the two survival curves compared in each plot ("NS" indicates no significant difference, "**" and "****" demonstrate significant difference with *p* < 0.01 and *p* < 0.0001). Dotted lines represent 95% confident intervals. In stage III samples, the median duration of follow-up for patients that died, until death was 458 days; for survivors this value

was 1247 days. The median total follow-up time for all stage III patients was 989 days. In stage IV samples, the median duration of follow-up for patients that died, until death was 357 days; for survivors this value was 1273 days. The median total follow-up time for all stage IV patients was 844 days.