High-throughput mass spectrometric N-glycomics

Reiding, K.R.

Chapter 6

Human plasma *N*-glycosylation as analyzed by MALDI-FTICR-MS associates with markers of inflammation and metabolic health

Research article

Karli R. Reiding[1], L. Renee Ruhaak[2], Hae-Won Uh[3], Said el Bouhaddani[3], Erik B. van den Akker[3,4], Rosina Plomp[1], Liam A. McDonnell[1], Jeanine J. Houwing-Duistermaat[3,5], P. Eline Slagboom[6], Marian Beekman[6], Manfred Wuhrer[1]

[1]Center for Proteomics and Metabolomics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands;

[2]Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, 2300 RC Leiden, The Netherlands;

[3]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands;

[4]Pattern Recognition & Bioinformatics, Delft University of Technology, 2600 GA Delft, The Netherlands;

[5]Department of Statistics, University of Leeds, LS2 9JT Leeds, United Kingdom;

[6]Department of Molecular Epidemiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands;

Reprinted and adapted with permission from Molecular & Cellular Protemics, December 8, 2016, DOI 10.1074/mcp.M116.065250

6.1 Abstract

Glycosylation is an abundant co- and post-translational protein modification of importance to protein processing and activity. Although not template-defined, glycosylation does reflect the biological state of an organism and is a high-potential biomarker for disease and patient stratification. However, to interpret a complex but informative sample like the total plasma *N*-glycome, it is important to establish its baseline association with plasma protein levels and systemic processes. Thus far, large-scale studies (n >200) of the total plasma *N*-glycome have been performed with methods of chromatographic and electrophoretic separation, which, although being informative, are limited in resolving the structural complexity of plasma *N*-glycans. MS has the opportunity to contribute additional information on, among others, antennarity, sialylation, and the identity of high-mannose type species.

Here, we have used matrix-assisted laser desorption/ionization (MALDI)-Fourier transform ion cyclotron resonance (FTICR)-MS to study the total plasma *N*-glycome of 2144 healthy middle-aged individuals from the Leiden Longevity Study, to allow association analysis with markers of metabolic health and inflammation. To achieve this, *N*-glycans were enzymatically released from their protein backbones, labeled at the reducing end with 2-aminobenzoic acid, and following purification analyzed by negative ion mode intermediate pressure MALDI-FTICR-MS. In doing so, we achieved the relative quantification of 61 glycan compositions, ranging from $Hex_4HexNAc_2$ to $Hex_7HexNAc_6dHex_1Neu5Ac_4$, as well as that of 39 glycosylation traits derived thereof. Next to confirming known associations of glycosylation with age and sex by MALDI-FTICR-MS, we report novel associations with C-reactive protein (CRP), interleukin 6 (IL-6), body mass index (BMI), leptin, adiponectin, HDL cholesterol, triglycerides (TG), insulin, gamma-glutamyl transferase (GGT), alanine aminotransferase (ALT), and smoking. Overall, the bisection, galactosylation, and sialylation of diantennary species, the sialylation of tetraantennary species, and the size of high-mannose species proved to be important plasma characteristics associated with inflammation and metabolic health.

6.2 Introduction

Glycosylation is a ubiquitous co- and post-translational protein modification of functional relevance to the processing and activity of the conjugate. Examples include quality control during protein folding, regulation of circulatory half-life, and modulation of receptor interactions by either providing the recognition motif or by affecting protein conformation[1-7]. Consequentially, glycosylation has been associated with a multitude of diseases and states thereof, among which the progression and metastasis of cancer and the remission of rheumatoid arthritis[8-11]. Because the process of glycosylation is not template-defined, glycosylation integrates a large series of cellular conditions such as glycosidase/glycosyltransferase abundance and activity, endoplasmic reticulum (ER)/Golgi localization and nucleotide sugar availability, and reflects the intricate biological state of an organism[1,2,12]. To establish glycosylation as biomarker for (early detection of) disease and patient stratification, analysis of an easily obtainable biofluid such as plasma is of great interest[13,14]. Observed effects in a total plasma *N*-glycome (TPNG), *i.e.*, the released *N*-glycans from all plasma proteins, are highly informative but difficult to comprehend because of the complex contributions from relative protein glycoforms and overall glycoprotein abundances[15,16].

To interpret the TPNG in a context of human health and disease it is of importance to establish the behavior of *N*-glycans, and groups of *N*-glycans, in relation to plasma protein levels and systemic processes such as inflammation and metabolism. Previous studies of suitable size (n > 200) have performed this to various degrees, finding plasma *N*-glycans to be highly associated with, *e.g.*, age, sex, inflammation, body mass index (BMI), cholesterol and lipid levels[17-24]. However, these studies have either been performed on single proteins, immunoglobulin G (IgG) being a particularly well-studied example, or predominantly by methods of liquid chromatographic and electrophoretic separation (*e.g.*, (ultra)-high-performance liquid chromatography (U)HPLC and capillary gel electrophoresis with laser-induced fluorescence detection (CGE-LIF)). Although of high analytical value—the techniques can separate analytes that are the same in monosaccharide composition (*e.g.*, $Hex_4HexNAc_4dHex_1$) but differ in glycan structure (*e.g.*, α1,3-branch galactosylation versus α1,6-branch galactosylation)—the complexity of the TPNG means that observed signals generally comprise a variety of distinct compositions[25-28]. Mass spectrometry provides orthogonal information from these methodologies, as it does not distinguish isomers but instead unambiguously evaluates glycans on a compositional level[29-31]. To date, a large mass spectrometric TPNG study remains to be performed for revealing associations with markers of inflammation and metabolic health[32-34].

Here, we have used high-resolution intermediate-pressure matrix-assisted laser desorption/ionization (MALDI)-Fourier transform ion cyclotron resonance (FTICR)-mass spectrometry (MS) to profile the total plasma *N*-glycosylation of 2144 middle-aged

individuals of the Leiden Longevity Study (LLS)[35]. Although MALDI-MS is reported to lead to underestimation of sialylated glycan species because of in-source and metastable decay, a phenomenon particularly visible with reflectron-based MALDI-time-of-flight (TOF)-MS, the intermediate pressure of the here-presented method prevents the residue loss and allows for the repeatable analysis of species carrying up to four sialic acids[36-41]. The 61 plasma *N*-glycan compositions we detected by the method, as well as 39 glycosylation traits mathematically derived thereof, showed to highly associate with not only age and sex, but also with clinical markers of inflammation, liver function, cholesterol, insulin, and lipid metabolism.

6.3 Experimental procedures

Participants

The LLS, described in detail previously[35,42], is a family-based study comprising 1671 offspring of 421 nonagenarians sibling pairs of Dutch descent, and the 744 partners of these offspring. A total of 2144 individuals with clinical blood parameters available were included in the current analysis. The study protocol was approved by the Leiden University Medical Center ethical committee and an informed consent was signed by all participants prior to participation in the study.

All standard blood measurements were performed in nonfasting venous blood samples using fully automated equipment. Glucose, high-sensitivity C-reactive protein (hsCRP), triglyceride (TG), total cholesterol, and high-density lipoprotein cholesterol (HDL) levels were measured on the Hitachi Modular P800 (Roche Diagnostics, Mannheim, Germany). Free triiodothyronine (T3) levels were measured on the Modular E170 (Roche Diagnostics). Low-density lipoprotein cholesterol (LDL) levels were calculated using the Friedewald formula[43], and set to missing if plasma TG levels exceeded 4.52 mmol/L. Insulin levels were measured on the Immulite 2500 (DPC, Los Angeles, CA). Specific sandwich enzyme-linked immunosorbent assays (ELISA) were u0sed for the determination of adiponectin (R&D Systems Europe, Abingdon, UK), leptin (Diagnostics Biochem Canada, Dorchester, Canada) and interleukin 6 (IL-6) levels (Sanquin Reagents, Amsterdam, The Netherlands). Alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels were measured using the NADH (with P-5'-P) methodology (Modular P800, Roche Diagnostics), and gamma-glutamyl transferase (GGT) levels using the L-gamma-glutamyl-3-carboxy-4-nitroanilide substrate methodology (Modular P800, Roche Diagnostics). Dehydroepiandrosterone sulfate (DHEA-S) levels were measured with an Architect delayed one-step immunoassay (Abbot, Wiesbaden, Germany). Hypertension was defined as having a systolic blood pressure > 140 and a diastolic blood pressure > 90. Antihypertensive medication included diuretics, beta-blockers, calcium channel blockers, and agents acting on the renin-

angiotensin system. Cytomegalovirus (CMV) serostatus was determined by ELISA using the CMV-IgG ELISA PKS assay (Medac, Wedel, Germany).

N-Glycan preparation

N-Glycans from total plasma proteins from participants of the LLS were released, labeled with 2-aminobenzoic acid (2-AA) (Sigma-Aldrich, Steinheim, Germany) to allow negative mode mass spectrometric detection, and purified using hydrophilic-interaction liquid chromatography (HILIC)-solid-phase extraction (SPE) as previously described[44]. Specifically, 20 µL of 2% sodium dodecyl sulfate (SDS) (US BioChem, Cleveland, OH) was added to 10 µL plasma, randomly distributed across twenty-seven 96-well plates, followed by protein denaturation for 10 min at 60 °C and subsequent neutralization of the SDS by 10 µL 4% Nonidet P-40 substitute (Nonidet P-40) (Sigma-Aldrich). Then, after addition of 0.5 mU peptide-*N*-glycosidase F (PNGase F; Roche Diagnostics) in 10 µL 5x phosphate buffered saline solution, the *N*-glycans were released overnight at 37 °C. Without intermediate purification, the *N*-glycans were labeled for 2 h at 65 °C with the addition of 50 µL 48 mg/mL 2-AA 63 mg/mL $NaCNBH_3$ (Merck, Darmstadt, Germany) in a 10:3 (*v/v*) mixture of dimethyl sulfoxide (DMSO; Sigma-Aldrich) and glacial acetic acid (Merck). HILIC-SPE was subsequently performed using 40 mg microcrystalline cellulose (Merck) in 96-well 0.45 µm GHP-filter plates (Pall, Ann Arbor, MI). All wells of the filter plate were washed using water and subsequently equilibrated using 80:20 (*v/v*) acetonitrile (ACN; Biosolve, Valkenswaard, The Netherlands)/water. The labeled *N*-glycan samples were then applied to the wells in 80% ACN, and the wells were washed using ACN/water (80:20 *v/v*). Purified 2-AA labeled *N*-glycans were eluted in 0.8-mL-deep well collection plates (ABgene via Westburg, Leusden, The Netherlands) using 400 µL water.

Carbon-SPE

Prior to analysis by MALDI-FTICR-MS, samples were additionally desalted using carbon SPE. To achieve this, 100 µL of graphitic porous carbon (Grace, Deerfield, IL) was applied to each well of an OF1100 96-well polypropylene filter plate with a 10 µm polyethylene frit (Orochem Technologies, Lombard, IL) using a 96-well column loader (Millipore, Billerica, MA). The stationary phase was activated and conditioned with 2 × 200 µL ACN/water (80:20 *v/v*) and 3 × 100 µL 0.1% trifluoroacetic acid (TFA; Sigma-Aldrich) in water, respectively. Of the 2-AA labeled *N*-glycans, 100 µL were loaded into the wells and washed using 3 × 100 µL 0.1% TFA in water. Slight vacuum was applied to facilitate the procedure. The 2-AA labeled *N*-glycans were eluted into a V-bottom microtiter plate (Nunc, Roskilde, Denmark) using 3 × 30 µL of freshly prepared ACN/water (80:20 *v/v*) containing 0.1% TFA by centrifugation at 500 rpm (154 mm rotational diameter).

MALDI-FTICR-MS analysis

One µL of 2-AA labeled *N*-glycans was spotted in quadruplicate on a 384-AnchorChip target plate (Bruker Daltonics, Bremen, Germany) and air-dried. Subsequently, 1 µL of 2,5-

dihydroxybenzoic acid (2,5-DHB; Bruker Daltonics) matrix (20 mg/mL in ACN/water; 50:50 (*v/v*)) was applied to the spots and left to dry. To generate microcrystals, 2 × 1 µL of ethanol was applied to the spots for recrystallization prior to mass spectrometric analysis.

The 9.4 T FTICR APEX-ultra mass spectrometer was equipped with a dual electrospray ionization (ESI)/MALDI ion source (Apollo II) incorporating a quadrupole mass filter and a smartbeam laser system. Before analysis, the instrument was externally calibrated by peptide calibration standard (Bruker Daltonics). All experiments used a laser spot size of ~150 µm, laser fluence slightly above threshold, and a laser repetition rate of 200 Hz. To allow the semiquantitative analysis for the range of expected *N*-glycans, all samples were analyzed using two methods: one optimized for lower *m/z* ions (approximately *m/z* 1000 to 2500) and another optimized for higher *m/z* ions (approximately *m/z* 2200 to 4000). The quadrupole was operated in rf-only mode with the selection masses set to *m/z* 1650 and 2500 for low-mass and high-mass measurements respectively. A customized experiment sequence (pulse program) was used, in which the multiple ICR-fill parameter was reconfigured to approximate a "random-walk" functionality[38]. Briefly, the ions produced from 50 laser shots were accumulated in a hexapole and then transferred through the rf-only quadrupole to the collision cell. The sample stage was then moved 200 µm, and fresh sample interrogated with the next 50 laser shots. This cycle was performed nine times, effectively accumulating ions from 450 laser shots in the collision cell. The accumulated ions were then transferred to the ICR cell for a mass analysis scan. Each spectrum is the sum of eight such scans. All data were acquired using ApexControl 3.0.0 software (Bruker Daltonics) in expert mode, controlled by Hystar 3.8 software (Bruker Daltonics) for automatic measurement. In total, 20,736 spectra were recorded, these being for each biological sample a quadruplicate of low- as well as high-mass measurements.

Data processing

Following acquisition, representative low- and high-mass spectra were internally calibrated in DataAnalysis 4.2 (Bruker Daltonics) using a set of expected glycan masses (**Supplemental Table S1**). Using the calibrated spectra, 37 glycan compositions were manually assigned within the low-mass measurements (H4N2 to H5N4S1), and 25 within the high-mass measurements (H5N4S1 to H7N6F1S4), using mass and parts-per-million (ppm) errors to validate the assignments (**Supplemental Table S1**) (H = hexose; N = *N*-acetylhexosamine; F = deoxyhexose (fucose); S = *N*-acetylneuraminic acid). In addition, peak widths were assessed per composition to allow precise area integration (**Supplemental Table S1**). The resulting 61 compositions (H5N4S1 being present in both the low- and high-mass spectra) were in agreement with previously reported observations, as well as knowledge of the biological synthesis of *N*-glycans[2,27,45,46].

To achieve repeated extraction of the list of glycan compositions from the 20,736 MALDI-FTICR-MS measurements (10,368 low-mass and 10,368 high-mass), the spectra were

converted to simple text based format (x,y) using msconvert from ProteoWizard 3.0.5622[47]. The raw mass spectrometric data has been made publicly available in the MassIVE repository (massive.ucsd.edu) titled "MALDI-FTICR-MS total plasma *N*-glycomics" with ID: MSV000080307. Spectrum calibration, analyte integration and spectrum curation was performed using MassyTools 0.1.5.0[48]. In short, spectra were calibrated by applying the least variance second degree polynomial fit through a set of calibration masses (**Supplemental Table S1**). During this step, spectra were excluded from further analysis when not all calibrants were detected with intensities at least 3-fold higher than the maximum signal deviation within the local noise (MinMax; roughly corresponding to a root-mean-square (RMS) signal-to-noise ratio (S/N) of 9). This led to the exclusion of 470 low-mass and 372 high-mass spectra. Glycan compositions from the analyte list were subsequently integrated by summing and grouping the areas of 95% of the theoretical isotopic envelope belonging to that composition. Prior to summation, each isotope was integrated using the peak widths previously established, and an equal width local background (within a window of 50 Thomson) was subtracted from these. To further ensure data quality, spectra were removed if more than 5% of the total analyte area was below S/N 3 (MinMax), which led to the additional exclusion of 305 low-mass and 424 high-mass spectra. After additional curation of clinical samples with no available information on age and sex, we retained a total of 16,346 spectra (8194 low-mass and 8152 high-mass), yielding glycosylation information on 2144 individuals.

To arrive at one set of glycan values per individual, the replicate spectra for the low- and high-mass spectra were averaged for that individual. In case of low-mass spectra, 1878 averages were constructed from 4 spectra, 170 from 3, 76 from 2 and 20 from 1 (**Supplemental Figure S1**). For the high-mass spectra, 1857 averages were constructed from 4 spectra, 177 from 3, 83 from 2 and 27 from 1. To reconstruct the overview of the total plasma *N*-glycome, the low- and high-mass averages were normalized on the value of their overlapping composition H5N4S1, and subsequently combined. The resulting combined pattern was normalized by dividing each glycan value by the sum of all glycan values (total area normalization). Hereof, we calculated derived glycosylation traits on basis of enzymatic steps and protein groupings (**Supplemental Table S2**). Glycan and derived trait variation within the replicate low-mass and high-mass measurements was assessed on the basis of mean, standard deviation (S.D.) and coefficient of variation (CV) (**Supplemental Figure S2**; **Supplemental Figure S3**). Mass spectrometric figures were exported from DataAnalysis 4.2 (Bruker Daltonics) and annotated with glycan depictions following the symbol nomenclature proposed by the Consortium for Functional Glycomics (CFG), created in GlycoWorkbench 2.1 build 146[49,50].

Data analysis

Throughout data analysis we employed R 3.1.2 in an environment of RStudio 0.98.1091 (RStudio Team, Boston, MA)[51]. In case of significance testing, a study-wide significance threshold was maintained of $\alpha = 1.00 \cdot 10^{-5}$, values below or equal being considered statistically significant. The value arises from being the lower bound of the order of magnitude of an $\alpha = 0.05$ significance threshold Bonferroni corrected for the total number of regression tests throughout the study (number of tests = 27 phenotypes · 100 glycan features + 26 sex comparisons + 25 age comparisons = 2751; $\alpha = 0.05/2751 = 1.82 \times 10^{-5} \approx 1.00 \times 10^{-5}$).

To limit the experimental component within the sample variability, batch correction was performed on the glycan and derived trait variables using the R package ComBat, using sample plate as batch[52]. To limit outlier influence, individual glycosylation values exceeding a 5 times S.D. value from the mean of that variable were excluded from statistical analysis. Insulin, hsCRP, IL-6, TG, adiponectin, leptin, ALT, AST, GGT, and DHEA-S levels were transformed to the natural logarithm because of nonnormal distribution of the data. In addition, to obtain interpretable estimates, the glycosylation variables were centered and scaled before analysis (*i.e.*, mean subtraction and division by S.D.).

Association of variables with age and sex

Linear and logistic regression analyses were performed to establish the association between age and sex (female = 0; male = 1) as outcome variables, and nonglycan clinical variables (**Table I**), glycans (**Table II**) and derived glycosylation traits (**Table III**) as predictor variables. As no glycosylation differences were found between LLS offspring and partners (a grouping to test predisposition for longevity), these individuals were grouped for all analyses. Furthermore, because the LLS contains multiple offspring from the same family, within-family (between-siblings) dependence was taken into account by using a sandwich estimator for the standard errors[53].

Association of glycosylation with clinical variables

To eliminate possible confounding effects, age, sex and the interaction thereof were included as covariates in further models. For these analyses, the remaining nonglycan variables were used as outcome (using linear and logistic regression for respectively continuous and dichotomous variables), whereas glycans and derived traits were used as predictor (model: nonglycan $\sim \beta_1 \cdot \text{age} + \beta_2 \cdot \text{sex} + \beta_3 \cdot \text{age*sex} + \beta_4 \cdot \text{glycan}$).

To visualize the association between nonglycans and (derived) glycan traits, the t-statistics (or Wald statistics in case of logistic regression, both $\beta_4/\text{S.E.}_4$) from the models were expressed in heatmap format. Sorting of the heatmap variables was performed using hierarchical clustering (Euclidean distance, complete linkage).

6.4 Results

To investigate the association of plasma protein *N*-glycosylation with clinical markers of metabolic health and inflammation, we analyzed the total plasma *N*-glycomes of 2144 middle-aged individuals of the LLS. *N*-Glycans were enzymatically released from their protein backbones, labeled at the reducing end with 2-AA, purified by HILIC- and carbon-SPE, and analyzed by intermediate pressure MALDI-FTICR-MS. The acidic tag 2-AA allowed the joint negative mode mass spectrometric detection and relative quantification of neutral and sialylated glycan species, whereas the intermediate pressure of the measurement limited the decay commonly observed for sialylated glycans with MALDI (**Figure 1**)[36-41].

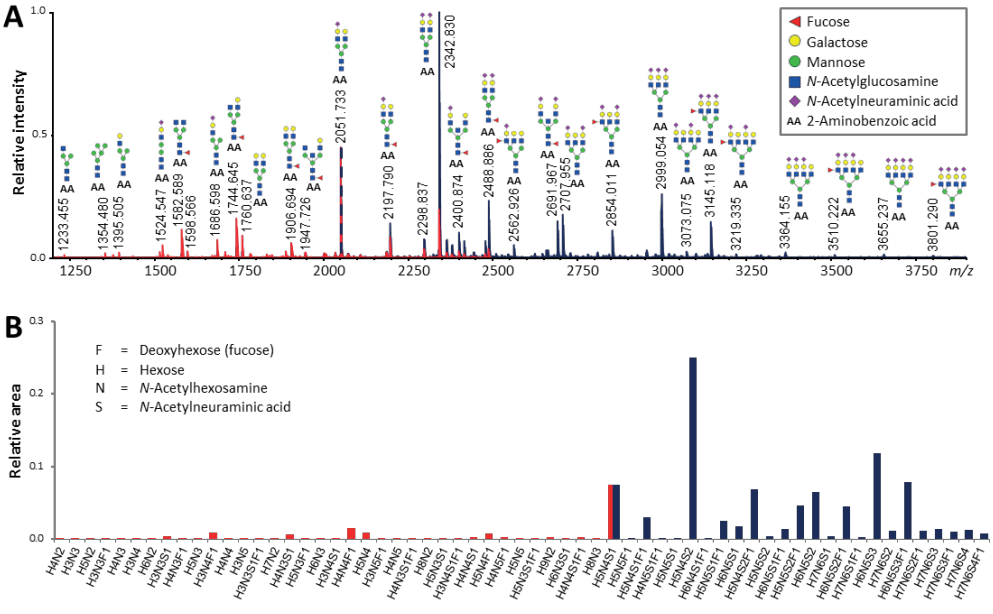

***Figure 1**. A typical total plasma N-glycome as analyzed by negative mode MALDI-FTICR-MS after enzymatic N-glycan release, 2-AA labeling, and purification. **A)** Combination of the low mass (red) and high mass (blue) mass spectra originating from a single case measurement. The relative abundances were normalized to the signal at m/z 2051.733, reflecting the N-glycan composition H5N4S1 [M-H]⁻. Whereas the glycan compositions could be established with high confidence, the displayed linkages are presumed based on literature knowledge[2,27,45,46]. **B)** Area integration of the detectable N-glycan compositions, summing the isotopes within 95% of the isotopic envelope for each species. Each composition is represented as a fraction of the total spectrum area.*

Measurement variability

Following spectrum curation, we retained a total of 16,346 mass spectra originating from low- and high-mass measurements of MALDI spotting quadruplicates for each individual. In the measurement optimized for lower masses (*m/z* 1000 to 2500) 37 *N*-glycans could be detected, ranging from H4N2 to H5N4S1, with an average absolute ppm error of 2.24 (S.D.

± 3.18). In the measurement optimized for higher masses (*m/z* 2200 to 4000) 25 additional *N*-glycans were detected, from the overlapping composition H5N4S1 to H7N6F1S4, the average absolute ppm error being 3.65 (S.D. ± 3.98) (**Supplemental Table S1**). Based on literature, the glycan compositions within the TPNG were presumed to have certain structural features[16,27,45,46]. Examples of this are the antennarity, judged as the number of *N*-acetylhexosamines minus two unless bisected, and bisection, judged to be the case if the number of *N*-acetylhexosamines equaled five and the number of hexoses five or less. Although these structural assignments are expected to represent the majority of structures contributing to an MS signal, additional structural isomers are likely to be present in the signals. For example, a composition assigned as tetraantennary may instead contain diantennary structures with two *N*-acetyllactosamine repeats, and the bisected species could be triantennary with incomplete galactosylation.

Assessing repeatability, an example quadruplicate measurement from a single individual yielded an average CV of 6.52% (S.D. ± 3.42%) for the 10 most abundant signals in the low-mass measurement (total area normalized for the mass range), and an average CV of 5.97% (S.D. ± 2.53%) for the 10 most abundant signals in the high-mass measurement (**Supplemental Figure S2A**; **Supplemental Figure S2B**). Combining the measurements by the overlapping composition H5N4S1 yielded for the 20 must abundant species an average CV of 9.29% (S.D. ± 5.88%) (**Supplemental Figure S2C**). Derived glycosylation traits, constructed to provide mathematical expressions of monosaccharide differences and groupings with structural similarity, showed a lower CV, a phenomenon previously observed for mass spectrometric plasma glycomics[54], *i.e.*, on average 1.20% (S.D. ± 0.84%) for the 20 most abundant members (**Supplemental Figure S2D**). For a listing of derived traits and their calculations see **Supplemental Table S2**.

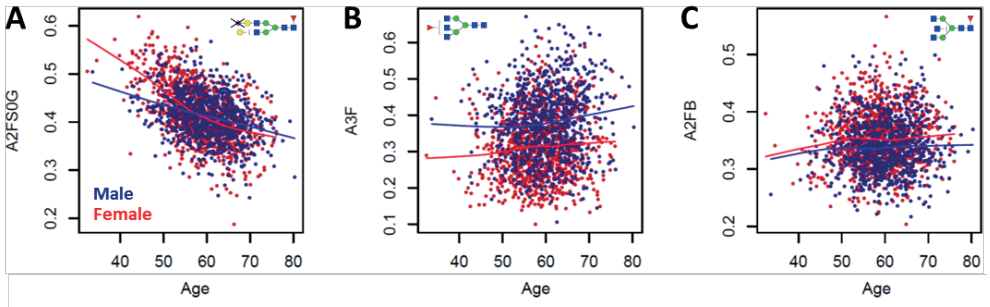

***Figure 2**. Scatterplots with local regression providing an overview of the three main derived glycosylation traits changing with age and differing by sex (males in blue, females in red). **A)** Galactosylation of nonsialylated fucosylated diantennary compositions (A2FS0G), to a large degree representative of IgG-Fc galactosylation[16]. **B)** Fucosylation of triantennary compositions (A3F). **C)** Bisection of diantennary fucosylated compositions (A2FB).*

In total, the glycosylation analysis workflow allowed for 2144 individuals the relative quantification of 61 *N*-glycan compositions and 39 derived traits. The LLS provided an additional 27 clinical variables to facilitate association analysis. Next to age and sex, measures were included on liver function (GGT, ALT, AST, AST/ALT), glucose metabolism (glucose, insulin, glucose/insulin), lipid metabolism (cholesterol, LDL-C, HDL-C, cholesterol/HDL-C, TG, lipid lowering medication, leptin, adiponectin), inflammation (hsCRP, IL-6), blood pressure (hypertension, antihypertensive medication), adrenal function (DHEA-S), thyroid function (free T3), as well as information on BMI, smoking and CMV infection, and familial propensity for longevity (**Table I**).

***Table I**. Non-glycan descriptives for the study population and association thereof with age and sex. Displayed are mean values with standard deviation (SD) for continuous variables, and the percentage of positive cases for binary variables. To assess if the variables differ by sex (female = 0; male = 1) and with age, respective logistic and linear regression was performed. Within-family dependence was taken into account by using a sandwich estimator for the standard errors. Effect sizes for the traits are displayed as coefficient for the trait (β) with standard error (SE). Displayed in bold are the p-values considered significant at or below the study-wide significance threshold of $\alpha = 1.0 \cdot 10^{-5}$.*

Phenotype	Total mean (SD) or % n = 2144	Female mean (SD) or % n = 1170	Male mean (SD) or % n = 974	Effect of trait with sex (F=0; M=1)		Effect of trait with age	
				β (SE)	*p*-value	β (SE)	*p*-value
Calendar age	59.2 (6.76)	58.6 (6.64)	59.8 (6.84)	0.19 (0.05)	6.1E-05	-	-
Alanine transamidase (IU/L)	24.2 (12.2)	21.8 (11.0)	27.1 (13.0)	**0.53 (0.05)**	**<2.2E-16**	-0.06 (0.14)	6.9E-01
Aspartate transamidase (IU/L)	27.0 (7.92)	25.9 (7.71)	28.3 (7.99)	**0.36 (0.05)**	**2.6E-13**	0.54 (0.14)	1.8E-04
AST ALT ratio	1.27 (0.53)	1.33 (0.52)	1.19 (0.54)	-0.33 (0.08)	5.7E-05	0.31 (0.15)	3.9E-02
Gamma-glutamyl transferase (IU/L)	31.4 (37.0)	25.7 (33.2)	38.3 (40.0)	**0.74 (0.06)**	**<2.2E-16**	0.47 (0.15)	2.2E-03
Non-fasted glucose (mmol/L)	5.87 (1.58)	5.75 (1.37)	6.01 (1.78)	0.17 (0.05)	2.4E-04	0.58 (0.15)	6.6E-05
Insulin (mU/L)	23.1 (22.0)	21.0 (17.8)	25.6 (25.9)	**0.21 (0.04)**	**1.3E-06**	0.54 (0.14)	1.8E-04
Glucose insulin ratio	0.46 (0.39)	0.49 (0.41)	0.43 (0.37)	-0.18 (0.04)	3.8E-05	-0.38 (0.14)	8.0E-03
Total cholesterol (mmol/L)	5.59 (1.17)	5.68 (1.21)	5.47 (1.12)	-0.19 (0.05)	3.4E-05	0.02 (0.16)	8.9E-01
Low-density lipoprotein cholesterol (mmol/L)	3.34 (0.96)	3.37 (0.98)	3.30 (0.94)	-0.07 (0.04)	1.1E-01	-0.10 (0.15)	5.0E-01
High-density lipoprotein cholesterol (mmol/L)	1.44 (0.46)	1.59 (0.47)	1.26 (0.36)	**-0.89 (0.06)**	**<2.2E-16**	-0.38 (0.15)	9.1E-03
Total cholesterol HDL ratio	4.20 (1.41)	3.82 (1.22)	4.65 (1.49)	**0.69 (0.05)**	**<2.2E-16**	0.40 (0.15)	7.0E-03
Triglycerides (mmol/L)	1.82 (1.16)	1.60 (0.93)	2.07 (1.34)	**0.48 (0.05)**	**<2.2E-16**	**0.74 (0.15)**	**5.3E-07**
Lipid lowering medication (%)	10.8%	8.87%	13.1%	0.14 (0.04)	1.6E-03	**1.13 (0.12)**	**<2.2E-16**
Leptin (ng/mL)	20.5 (21.6)	29.3 (24.7)	9.92 (9.83)	**-1.61 (0.09)**	**<2.2E-16**	0.17 (0.15)	2.6E-01
Adiponectin (mg/L)	6.27 (3.30)	7.49 (3.55)	4.81 (2.23)	**-1.04 (0.06)**	**<2.2E-16**	0.08 (0.15)	5.7E-01
Body mass index	25.4 (3.58)	25.1 (4.00)	25.7 (2.97)	0.18 (0.05)	7.1E-04	0.41 (0.15)	4.4E-03
Hypertension (%)	24.2%	24.3%	24.0%	-0.01 (0.05)	8.8E-01	**1.12 (0.15)**	**5.0E-14**
Antihypertensive medication (%)	20.5%	20.3%	20.7%	0.01 (0.05)	7.5E-01	**1.38 (0.15)**	**<2.2E-16**
Dehydroepiandrosterone sulfate (µmol/L)	4.45 (2.80)	3.59 (2.21)	5.49 (3.07)	**0.85 (0.06)**	**<2.2E-16**	**-1.87 (0.16)**	**<2.2E-16**
High sensitivity C-reactive protein (mg/L)	3.15 (11.0)	3.11 (10.4)	3.21 (11.7)	-0.05 (0.04)	2.9E-01	0.63 (0.15)	3.5E-05
Interleukin 6 (pg/mL)	0.68 (1.46)	0.62 (1.17)	0.75 (1.74)	0.11 (0.04)	9.8E-03	**0.77 (0.15)**	**5.6E-07**
Smoking (%)	13.6%	13.2%	14.1%	0.02 (0.05)	6.1E-01	**-0.90 (0.15)**	**2.4E-09**
Free triiodothyronine (pmol/L)	4.11 (0.72)	3.94 (0.71)	4.31 (0.67)	**0.62 (0.07)**	**<2.2E-16**	-0.55 (0.14)	9.4E-05
Cytomegalovirus infection (%)	46.9%	50.6%	42.3%	-0.17 (0.05)	5.3E-04	0.56 (0.16)	4.9E-04
Member of long-lived family (%)	69.2%	67.4%	71.4%	0.09 (0.04)	5.3E-02	0.33 (0.17)	5.0E-02

***Table II**. N-glycan descriptives and their association with age and sex. Displayed are mean values with SD. To assess if the variables differ by sex (female = 0; male = 1) and with age, respective logistic and linear regression was performed, adjusted for within-family dependence. Effect sizes for the traits are displayed as coefficient for the trait (β) with SE, all of which are representative of a 1 SD increase in the glycosylation value. Displayed in bold are the p-values considered significant at or below the study-wide significance threshold of $\alpha = 1.0 \cdot 10^{-5}$.*

Composition	*m/z* [M-H]-	Depiction	Total mean % (SD)	Female mean % (SD)	Male mean % (SD)	Effect of trait with sex (F=0; M=1)		Effect of trait with age	
						β (SE)	*p*-value	β (SE)	*p*-value
H4N2	1192.43		0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	-0.04 (0.04)	2.9E-01	-0.19 (0.13)	1.6E-01
H3N3	1233.45		0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.09 (0.05)	4.1E-02	-0.45 (0.15)	3.5E-03
H5N2	1354.48		0.14 (0.04)	0.14 (0.04)	0.13 (0.04)	-0.10 (0.04)	2.4E-02	-0.22 (0.14)	1.2E-01
H3N3F1	1379.51		0.06 (0.02)	0.06 (0.02)	0.06 (0.02)	0.15 (0.04)	2.2E-04	0.20 (0.15)	1.8E-01
H4N3	1395.51		0.11 (0.03)	0.11 (0.03)	0.11 (0.03)	-0.02 (0.04)	5.5E-01	-0.13 (0.13)	3.1E-01
H3N4	1436.53		0.06 (0.03)	0.06 (0.03)	0.06 (0.03)	0.11 (0.05)	1.7E-02	**1.21 (0.14)**	**<2.2E-16**
H6N2	1516.53		0.20 (0.06)	0.20 (0.06)	0.20 (0.06)	-0.06 (0.04)	1.5E-01	-0.54 (0.14)	2.1E-04
H3N3S1	1524.55		0.34 (0.07)	0.34 (0.07)	0.34 (0.07)	0.03 (0.04)	4.6E-01	-0.23 (0.14)	9.0E-02
H4N3F1	1541.56		0.06 (0.02)	0.06 (0.02)	0.06 (0.02)	0.13 (0.04)	1.8E-03	**-0.97 (0.15)**	**1.1E-10**
H5N3	1557.56		0.04 (0.01)	0.04 (0.01)	0.03 (0.01)	-0.16 (0.05)	4.5E-04	-0.44 (0.14)	2.6E-03
H3N4F1	1582.59		1.18 (0.49)	1.15 (0.49)	1.23 (0.48)	0.18 (0.04)	3.1E-05	**1.50 (0.14)**	**<2.2E-16**
H4N4	1598.58		0.19 (0.06)	0.19 (0.06)	0.19 (0.06)	0.00 (0.04)	9.5E-01	-0.10 (0.15)	5.0E-01
H3N5	1639.61		0.06 (0.02)	0.06 (0.02)	0.06 (0.02)	0.10 (0.04)	2.9E-02	**1.68 (0.14)**	**<2.2E-16**
H3N3F1S1	1670.61		0.05 (0.01)	0.05 (0.01)	0.06 (0.01)	**0.21 (0.04)**	**8.3E-07**	-0.30 (0.15)	4.4E-02
H7N2	1678.58		0.08 (0.02)	0.08 (0.02)	0.07 (0.02)	-0.03 (0.04)	5.5E-01	-0.58 (0.15)	8.0E-05
H4N3S1	1686.60		0.61 (0.12)	0.60 (0.12)	0.61 (0.12)	0.08 (0.04)	4.8E-02	-0.12 (0.14)	4.0E-01
H5N3F1	1703.62		0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	-0.08 (0.04)	8.3E-02	-0.53 (0.14)	1.9E-04
H6N3	1719.61		0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	0.06 (0.04)	1.7E-01	**0.72 (0.15)**	**1.7E-06**
H3N4S1	1727.63		0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	-0.03 (0.04)	4.6E-01	-0.26 (0.14)	5.6E-02
H4N4F1	1744.64		1.40 (0.46)	1.36 (0.45)	1.44 (0.47)	**0.19 (0.04)**	**9.1E-06**	-0.33 (0.15)	3.1E-02
H5N4	1760.64		0.90 (0.18)	0.90 (0.18)	0.89 (0.18)	-0.06 (0.05)	1.6E-01	-0.32 (0.14)	2.4E-02
H3N5F1	1785.67		0.30 (0.11)	0.29 (0.11)	0.30 (0.11)	0.09 (0.04)	4.8E-02	**1.67 (0.14)**	**<2.2E-16**
H4N5	1801.66		0.07 (0.02)	0.06 (0.02)	0.07 (0.02)	0.05 (0.05)	2.7E-01	**0.83 (0.15)**	**1.6E-08**
H4N3F1S1	1832.66		0.10 (0.02)	0.10 (0.02)	0.10 (0.02)	0.15 (0.04)	5.0E-04	-0.46 (0.15)	2.3E-03
H8N2	1840.64		0.12 (0.03)	0.12 (0.03)	0.12 (0.03)	-0.12 (0.04)	4.2E-03	**-0.73 (0.14)**	**4.5E-07**
H5N3S1	1848.65		0.10 (0.02)	0.11 (0.02)	0.10 (0.02)	-0.20 (0.05)	1.6E-05	**0.73 (0.15)**	**1.1E-06**
H3N4F1S1	1873.69		0.03 (0.01)	0.03 (0.01)	0.03 (0.01)	-0.02 (0.04)	7.3E-01	0.13 (0.14)	3.7E-01
H4N4S1	1889.68		0.37 (0.08)	0.37 (0.09)	0.36 (0.08)	**-0.19 (0.04)**	**9.0E-06**	0.25 (0.14)	6.9E-02
H5N4F1	1906.70		0.62 (0.19)	0.61 (0.20)	0.63 (0.19)	0.14 (0.04)	1.4E-03	**-1.60 (0.15)**	**<2.2E-16**
H4N5F1	1947.72		0.29 (0.09)	0.28 (0.09)	0.29 (0.10)	0.06 (0.04)	1.6E-01	0.09 (0.15)	5.5E-01
H5N5	1963.72		0.03 (0.01)	0.03 (0.01)	0.03 (0.01)	-0.12 (0.04)	8.7E-03	-0.43 (0.15)	5.3E-03

Association of glycosylation with age and sex

Glycosylation was found to highly associate with age and sex by respectively linear and logistic regression analysis (**Figure 2**; **Table II**; **Table III**). A GEE approach was used for all statistical analyses to adjust the standard errors (S.E.) for between-sibling dependence, and a study-wide significance threshold was maintained of $\alpha = 1.0 \cdot 10^{-5}$. Changes with aging included a decrease of galactosylation of diantennary glycans, visible most specifically for the galactosylation of nonsialylated diantennaries with fucose ($\beta_{A2FS0G} = -2.82$ S.E. $\pm$ 0.13; $p_{A2FS0G} < 2.2 \times 10^{-16}$) and without fucose ($\beta_{A2F0S0G} = -1.34 \pm 0.15$; $p_{A2F0S0G} < 2.2 \times 10^{-16}$). These changes were mainly driven by the increases in glycan compositions H3N4 ($\beta_{H3N4} = 1.21 \pm 0.14$; $p_{H3N4} < 2.2 \times 10^{-16}$), H3N5 ($\beta_{H3N5} = 1.68 \pm 0.14$; $p_{H3N5} < 2.2 \times 10^{-16}$), H3N4F1 ($\beta_{H3N4F1} =$

1.50 ± 0.14; $p_{H3N4F1} < 2.2 \times 10^{-16}$), H3N5F1 ($\beta_{H3N5F1} = 1.67 \pm 0.14$; $p_{H3N5F1} < 2.2 \times 10^{-16}$) and the decreases in H5N4F1 ($\beta_{H5N4F1} = -1.60 \pm 0.15$; $p_{H5N4F1} < 2.2 \times 10^{-16}$) and H5N4F1S1 ($\beta_{H5N4F1S1} = -0.81 \pm 0.15$; $p_{H5N4F1S1} = 4.5 \times 10^{-8}$). Further increasing with age were the bisection of nonsialylated fucosylated diantennaries ($\beta_{A2FS0B} = 0.77 \pm 0.14$; $p_{A2FS0B} = 6.7 \times 10^{-8}$), sialylation per galactose of fucosylated diantennaries ($\beta_{A2FGS} = 1.14 \pm 0.15$; $p_{A2FGS} = 9.8 \times 10^{-15}$), and the fucosylation of both tri- and tetraantennary compositions ($\beta_{A3F} = 0.82 \pm 0.15$; $p_{A3F} = 7.4 \times 10^{-8}$ and $\beta_{A4F} = 0.77 \pm 0.15$; $p_{A4F} = 2.0 \times 10^{-7}$).

Table II *continued.*

Composition	*m/z* [M-H]-	Depiction	Total mean % (SD)	Female mean % (SD)	Male mean % (SD)	Effect of trait with sex (F=0; M=1)		Effect of trait with age	
						β (SE)	*p*-value	β (SE)	*p*-value
H5N3F1S1	1994.71		0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.04)	9.9E-01	-0.33 (0.15)	2.6E-02
H9N2	2002.69		0.20 (0.04)	0.20 (0.04)	0.20 (0.04)	0.12 (0.04)	6.2E-03	-0.21 (0.14)	1.4E-01
H6N3S1	2010.71		0.09 (0.02)	0.09 (0.02)	0.09 (0.02)	-0.16 (0.04)	4.9E-04	0.29 (0.14)	4.5E-02
H4N4F1S1	2035.74		0.23 (0.05)	0.23 (0.05)	0.24 (0.05)	**0.20 (0.04)**	**2.6E-06**	-0.31 (0.15)	3.6E-02
H8N3	2043.72		0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02 (0.04)	6.4E-01	**-0.84 (0.14)**	**5.0E-09**
H5N4S1	2051.73		6.94 (0.88)	6.97 (0.89)	6.91 (0.87)	-0.06 (0.04)	1.5E-01	0.01 (0.14)	9.4E-01
H5N5F1	2109.77		0.15 (0.05)	0.15 (0.05)	0.15 (0.05)	-0.04 (0.04)	3.4E-01	-0.62 (0.15)	4.7E-05
H5N4F1S1	2197.79		2.14 (0.46)	2.08 (0.45)	2.20 (0.46)	**0.27 (0.05)**	**5.1E-09**	**-0.81 (0.15)**	**4.5E-08**
H4N5F1S1	2238.82		0.10 (0.04)	0.10 (0.04)	0.10 (0.04)	-0.01 (0.04)	7.2E-01	**0.95 (0.14)**	**3.4E-12**
H5N5S1	2254.81		0.16 (0.05)	0.16 (0.05)	0.16 (0.05)	-0.03 (0.04)	5.1E-01	-0.20 (0.15)	1.7E-01
H5N4S2	2342.83		24.03 (2.94)	23.9 (2.93)	24.2 (2.94)	0.09 (0.04)	2.5E-02	-0.33 (0.14)	1.7E-02
H6N4F1S1	2359.84		0.13 (0.04)	0.13 (0.04)	0.13 (0.05)	0.04 (0.04)	3.2E-01	-0.29 (0.15)	5.3E-02
H5N5F1S1	2400.87		1.74 (0.55)	1.77 (0.57)	1.72 (0.52)	-0.10 (0.04)	1.6E-02	-0.26 (0.15)	9.2E-02
H6N5S1	2416.87		1.55 (0.30)	1.64 (0.30)	1.45 (0.28)	**-0.71 (0.05)**	**<2.2E-16**	-0.51 (0.14)	3.4E-04
H5N4F1S2	2488.89		4.38 (0.92)	4.20 (0.84)	4.59 (0.95)	**0.45 (0.05)**	**<2.2E-16**	0.05 (0.14)	7.2E-01
H5N5S2	2545.91		0.26 (0.06)	0.27 (0.06)	0.25 (0.06)	-0.18 (0.04)	3.2E-05	-0.54 (0.14)	1.6E-04
H6N5F1S1	2562.92		0.93 (0.23)	0.88 (0.21)	0.99 (0.23)	**0.53 (0.05)**	**<2.2E-16**	**0.71 (0.15)**	**1.1E-06**
H5N5F1S2	2691.97		2.89 (0.87)	2.84 (0.85)	2.96 (0.88)	0.14 (0.04)	1.9E-03	0.24 (0.15)	1.1E-01
H6N5S2	2707.96		7.65 (1.54)	8.08 (1.49)	7.13 (1.43)	**-0.70 (0.05)**	**<2.2E-16**	-0.59 (0.15)	6.6E-05
H7N6S1	2782.00		0.50 (0.16)	0.54 (0.16)	0.46 (0.14)	**-0.58 (0.05)**	**<2.2E-16**	0.07 (0.13)	6.1E-01
H6N5F1S2	2854.02		4.02 (1.20)	3.73 (1.09)	4.38 (1.22)	**0.58 (0.05)**	**<2.2E-16**	**0.77 (0.15)**	**2.2E-07**
H7N6F1S1	2928.06		0.28 (0.10)	0.27 (0.10)	0.30 (0.10)	**0.38 (0.05)**	**<2.2E-16**	**0.87 (0.14)**	**1.8E-09**
H6N5S3	2999.06		14.99 (3.25)	15.8 (3.17)	14.0 (3.07)	**-0.61 (0.05)**	**<2.2E-16**	-0.67 (0.16)	1.6E-05
H7N6S2	3073.09		1.74 (0.47)	1.85 (0.48)	1.62 (0.43)	**-0.53 (0.05)**	**<2.2E-16**	0.05 (0.14)	7.2E-01
H6N5F1S3	3145.11		7.81 (2.53)	7.15 (2.31)	8.59 (2.57)	**0.61 (0.05)**	**<2.2E-16**	**0.74 (0.15)**	**8.1E-07**
H7N6F1S2	3219.15		1.08 (0.35)	1.01 (0.33)	1.17 (0.35)	**0.48 (0.05)**	**<2.2E-16**	**0.90 (0.14)**	**2.4E-10**
H7N6S3	3364.19		2.35 (0.62)	2.48 (0.62)	2.20 (0.57)	**-0.50 (0.05)**	**<2.2E-16**	-0.07 (0.15)	6.3E-01
H7N6F1S3	3510.25		1.20 (0.36)	1.12 (0.34)	1.28 (0.36)	**0.46 (0.05)**	**<2.2E-16**	**0.88 (0.15)**	**1.2E-09**
H7N6S4	3655.28		1.96 (0.47)	2.01 (0.48)	1.89 (0.46)	**-0.26 (0.05)**	**5.8E-08**	0.16 (0.15)	2.8E-01
H7N6F1S4	3801.34		1.05 (0.30)	0.99 (0.28)	1.12 (0.31)	**0.45 (0.05)**	**<2.2E-16**	**0.79 (0.14)**	**1.8E-08**

***Table III**. Derived glycosylation trait descriptives and their association with age and sex. Displayed are mean values with SD. To assess if the variables differ by sex (female = 0; male = 1) and with age, respective logistic and linear regression was performed, adjusted for within-family dependence. Effect sizes for the traits are displayed as coefficient for the trait (β) with SE, and are representative of a 1 SD increase in the glycosylation value. Displayed in bold are the p-values considered significant at or below the study-wide significance threshold of $\alpha = 1.0 \cdot 10^{-5}$.*

Derived trait	Description	Total mean % (SD)	Female mean % (SD)	Male mean % (SD)	Effect of trait with sex (F=0; M=1)		Effect of trait with age	
					β (SE)	*p*-value	β (SE)	*p*-value
Complexity								
M	Overall high mannose type	0.78 (0.17)	0.78 (0.17)	0.78 (0.17)	-0.03 (0.04)	4.5E-01	-0.51 (0.14)	3.9E-04
MM	High mannose occupancy	690 (11.0)	689 (11.0)	691 (11.0)	0.19 (0.04)	2.1E-05	-0.12 (0.14)	4.0E-01
Hy	Overall hybrid type	0.44 (0.09)	0.45 (0.09)	0.44 (0.09)	-0.07 (0.04)	1.0E-01	0.06 (0.14)	6.6E-01
C	Overall complex type	97.3 (0.46)	97.3 (0.47)	97.3 (0.45)	-0.02 (0.04)	7.0E-01	0.36 (0.14)	1.0E-02
A1	Overall monoantennary	1.16 (0.22)	1.15 (0.23)	1.17 (0.22)	0.10 (0.04)	1.6E-02	-0.21 (0.14)	1.3E-01
A2	Overall diantennary	49.2 (5.32)	48.8 (5.30)	49.7 (5.30)	0.18 (0.04)	3.5E-05	-0.09 (0.15)	5.5E-01
A3	Overall triantennary	37.1 (4.50)	37.5 (4.46)	36.7 (4.51)	-0.18 (0.04)	2.2E-05	-0.10 (0.15)	5.0E-01
A4	Overall tetraantennary	10.2 (1.95)	10.3 (2.01)	10.1 (1.86)	-0.13 (0.04)	3.4E-03	0.52 (0.14)	2.6E-04
Fucosylation								
F	Overall	33.0 (5.58)	31.4 (5.21)	35.0 (5.37)	**0.72 (0.05)**	**<2.2E-16**	**0.78 (0.15)**	**3.9E-07**
A2F	Within A2	32.4 (4.47)	31.9 (4.48)	32.9 (4.41)	**0.22 (0.04)**	**3.3E-07**	0.19 (0.15)	1.9E-01
A3F	Within A3	34.7 (9.87)	31.7 (9.04)	38.2 (9.67)	**0.73 (0.05)**	**<2.2E-16**	**0.82 (0.15)**	**7.4E-08**
A4F	Within A4	35.6 (8.41)	33.1 (7.85)	38.7 (8.05)	**0.74 (0.05)**	**<2.2E-16**	**0.77 (0.15)**	**2.0E-07**
Bisection								
B	Overall	6.12 (1.52)	6.09 (1.54)	6.17 (1.51)	0.05 (0.04)	2.2E-01	0.18 (0.15)	2.2E-01
A2B	Within A2	12.5 (2.79)	12.5 (2.86)	12.4 (2.71)	-0.05 (0.04)	2.1E-01	0.28 (0.15)	5.8E-02
A2F0B	Within nonfucosylated A2	1.77 (0.41)	1.78 (0.41)	1.75 (0.41)	-0.09 (0.04)	3.7E-02	0.16 (0.15)	3.1E-01
A2FB	Within fucosylated A2	34.7 (4.87)	35.3 (4.84)	34.0 (4.81)	**-0.27 (0.04)**	**3.8E-10**	0.24 (0.14)	8.7E-02
A2FS0B	Within nonsialylated fucosylated A2	18.9 (2.99)	19.2 (2.94)	18.5 (3.01)	**-0.25 (0.04)**	**1.1E-08**	**0.77 (0.14)**	**6.7E-08**
A2FSB	Within sialylated fucosylated A2	40.9 (5.91)	41.6 (5.88)	40.1 (5.85)	**-0.25 (0.04)**	**4.5E-09**	0.24 (0.15)	9.4E-02
Galactosylation per antenna								
AG	Overall	96.5 (1.11)	96.6 (1.10)	96.4 (1.11)	-0.15 (0.04)	3.9E-04	**-0.94 (0.15)**	**1.3E-10**
A2G	Within A2	93.3 (1.68)	93.4 (1.71)	93.2 (1.65)	-0.10 (0.04)	1.5E-02	**-1.25 (0.14)**	**<2.2E-16**
A2F0G	Within nonfucosylated A2	98.5 (0.30)	98.5 (0.30)	98.5 (0.29)	0.05 (0.04)	2.6E-01	**-0.99 (0.14)**	**5.4E-13**
A2F0S0G	Within nonsialylated nonfucosylated A2	81.1 (3.84)	81.3 (3.94)	80.8 (3.69)	-0.14 (0.05)	2.4E-03	**-1.32 (0.15)**	**<2.2E-16**
A2FG	Within fucosylated A2	82.6 (3.97)	82.6 (4.04)	82.6 (3.89)	-0.03 (0.04)	4.8E-01	**-1.36 (0.14)**	**<2.2E-16**
A2FS0G	Within nonsialylated fucosylated A2	41.4 (5.40)	41.7 (5.80)	41.2 (4.86)	-0.10 (0.04)	2.1E-02	**-2.82 (0.13)**	**<2.2E-16**
A2FSG	Within sialylated fucosylated A2	98.5 (0.26)	98.5 (0.26)	98.5 (0.26)	0.10 (0.04)	1.7E-02	-0.50 (0.14)	3.5E-04
Sialylation per antenna								
AS	Overall	79.1 (1.71)	79.1 (1.69)	79.1 (1.72)	0.02 (0.04)	5.5E-01	-0.41 (0.15)	5.2E-03
A2S	Within A2	76.6 (2.47)	76.6 (2.45)	76.6 (2.50)	0.00 (0.04)	9.8E-01	-0.35 (0.15)	1.6E-02
A2F0S	Within nonfucosylated A2	84.5 (1.31)	84.4 (1.30)	84.6 (1.32)	0.18 (0.04)	1.0E-05	-0.56 (0.13)	3.4E-05
A2FS	Within fucosylated A2	60.1 (4.79)	60.0 (4.71)	60.3 (4.89)	0.05 (0.04)	2.3E-01	-0.20 (0.15)	1.7E-01
Sialylation per galactose								
GS	Overall	81.9 (1.18)	81.8 (1.18)	82.0 (1.17)	0.17 (0.04)	6.0E-05	0.23 (0.15)	1.2E-01
A2GS	Within A2	82.1 (1.53)	82.0 (1.53)	82.2 (1.52)	0.11 (0.04)	7.4E-03	0.63 (0.15)	2.4E-05
A2F0GS	Within nonfucosylated A2	85.8 (1.17)	85.7 (1.16)	85.8 (1.18)	0.19 (0.04)	1.2E-05	-0.34 (0.14)	1.4E-02
A2FGS	Within fucosylated A2	72.7 (3.37)	72.5 (3.38)	73.0 (3.33)	0.14 (0.04)	1.2E-03	**1.14 (0.15)**	**9.8E-15**
A3GS	Within A3	84.8 (1.01)	84.7 (0.98)	84.8 (1.05)	0.12 (0.04)	3.5E-03	-0.15 (0.14)	2.8E-01
A3F0GS	Within nonfucosylated A3	84.9 (1.09)	84.9 (1.05)	84.9 (1.14)	0.01 (0.04)	7.8E-01	-0.30 (0.14)	3.7E-02
A3FGS	Within fucosylated A3	84.2 (1.32)	84.0 (1.32)	84.5 (1.28)	**0.34 (0.04)**	**2.7E-14**	0.22 (0.14)	1.2E-01
A4GS	Within A4	71.4 (1.93)	71.3 (1.88)	71.6 (1.96)	0.16 (0.04)	9.0E-05	-0.11 (0.14)	4.3E-01
A4F0GS	Within nonfucosylated A4	71.9 (2.16)	71.6 (2.05)	72.3 (2.22)	**0.34 (0.04)**	**2.5E-14**	0.09 (0.14)	5.1E-01
A4FGS	Within fucosylated A4	70.8 (1.96)	70.8 (1.95)	70.7 (1.97)	-0.08 (0.04)	5.2E-02	-0.24 (0.14)	8.7E-02

Fucosylation of triantennary and tetraantennary structures (A3F and A4F) proved also to be a major glycosylation difference between females and males (female = 0; male = 1) (β_{A3F} =0.73 ± 0.05; $p_{A3F} < 2.2 \times 10^{-16}$ and $\beta_{A4F} = 0.74 \pm 0.05$; $p_{A4F} < 2.2 \times 10^{-16}$), driven by higher male values in all fucosylated tri- and tetraantennary compositions, such as H6N5F1S3 ($\beta_{H6N5F1S3} = 0.61 \pm 0.05$; $p_{H6N5F1S3} < 2.2 \times 10^{-16}$) and H7N6F1S4 ($\beta_{H7N6F1S4} = 0.45 \pm 0.05$; $p_{H7N6F1S4} < 2.2 \times 10^{-16}$), and significantly lower levels of all nonfucosylated tri- and tetraantennary compositions, such as H6N5S3 (β_{H6N5S3} =−0.61 ± 0.05; $p_{H6N5S3} < 2.2 \times 10^{-16}$) and H7N6S4 ($\beta_{H7N6S4}$ = −0.26 ± 0.05; $p_{H7N6S4} = 5.8 \times 10^{-8}$). Furthermore, males proved to have lower bisection of diantennary fucosylated species (β_{A2FB} = −0.27 ± 0.04; $p = 3.8 \times 10^{-10}$) when compared with females, but did have a higher sialylation per galactose of tetraantennary nonfucosylated compositions ($\beta_{A4F0GS} = 0.34 \pm 0.04$; $p = 2.5 \times 10^{-14}$).

Association glycosylation with inflammation and metabolic health

Regression analysis was used to establish the relationship between the glycosylation traits and clinical markers, adding age, sex and the interaction thereof as covariates to limit their confounding influence (**Supplemental Table S3**). The t-statistics (or Wald-statistics; both β_4/S.E.$_4$) arising from the models were expressed in clustered heatmap format (**Figure 3**; for a heatmap visualization of the results without age and sex adjustment see **Supplemental Figure S4**). Significantly associating with a selection of derived glycosylation traits were hsCRP (23 statistically significant associations out of a possible 39), GGT (14), BMI (13), leptin (13), smoking (10), TG (9), insulin (7), the ratio of total cholesterol and HDL (7), the ratio of glucose and insulin (5), adiponectin (5), HDL (4), IL-6 (3), ALT (2), the ratio of ASL and ALT (1) and lipid medication (1) (**Supplemental Table S4**; **Supplemental Table S5**). Only individual *N*-glycan associations could be proven for AST, glucose, cholesterol, LDL and free T3, whereas no associations were found for hypertension, the usage of antihypertensive medication, DHEA-S and CMV infection.

Inflammatory marker hsCRP showed the most associations with the total plasma *N*-glycome, including a positive association with tri- and tetraantennary glycans ($\beta_{A3} = 0.24 \pm 0.03$; $p_{A3} < 2.2 \times 10^{-16}$ and $\beta_{A4} = 0.20 \pm 0.03$; $p_{A4} = 2.9 \times 10^{-13}$) at the expense of high-mannose (β_M = −0.18 ± 0.02; $p_M = 1.2 \times 10^{-13}$), hybrid ($\beta_{Hy}$ = −0.24 ± 0.03; $p_{Hy} < 2.2 \times 10^{-16}$), monoantennary ($\beta_{A1}$ = −0.17 ± 0.02; $p_{A1} = 1.8 \times 10^{-12}$) and diantennary species (β_{A2} = −0.25 ± 0.03; $p_{A2} < 2.2 \times 10^{-16}$). Although fucosylation of diantennary species proved to decrease with higher hsCRP levels (β_{A2F} = −0.13 ± 0.02; $p_{A2F} = 1.5 \times 10^{-7}$), an increase was seen in the fucosylation of triantennary species (β_{A3F} = 0.13 ± 0.03; $p_{A3F} = 1.0 \times 10^{-6}$). Additional increases were found for sialylation of (most specifically) fucosylated diantennary (β_{A2FGS} = 0.25 ± 0.02; $p_{A2FGS} < 2.2 \times 10^{-16}$) and triantennary ($\beta_{A3FGS} = 0.15 \pm 0.02$; $p_{A3FGS} = 1.0 \times 10^{-9}$) species, as well as an increase in average high-mannose size ($\beta_{MM} = 0.13 \pm 0.02$; $p_{MM} = 7.9 \times 10^{-8}$) and a decrease in bisection of the nonfucosylated diantennaries in particular (β_{A2F0B} = −0.14 ± 0.02; $p_{A2F0B} = 1.4 \times 10^{-8}$). Notably, although galactosylation of nonsialylated

diantennaries without fucose increased ($\beta_{A2F0S0G}$ = 0.12 ± 0.02; $p_{A2F0S0G}$ = 2.6 × 10^{-7}), galactosylation of the same species, but with fucose, decreased instead (β_{A2FS0G} = −0.20 ± 0.03; p_{A2FS0G} = 1.7 × 10^{-12}). Interestingly, the associations observed with hsCRP could only in part be translated to the upstream cytokine IL-6[55,56], which only showed reproduction of the decreased A2F galactosylation (β_{A2FS0G} = −0.19 ± 0.03; p_{A2FS0G} = 4.6 × 10^{-12}) and increased sialylation per galactose thereof (β_{A2FGS} = 0.25 ± 0.02; p_{A2FGS} < 2.2 × 10^{-16}).

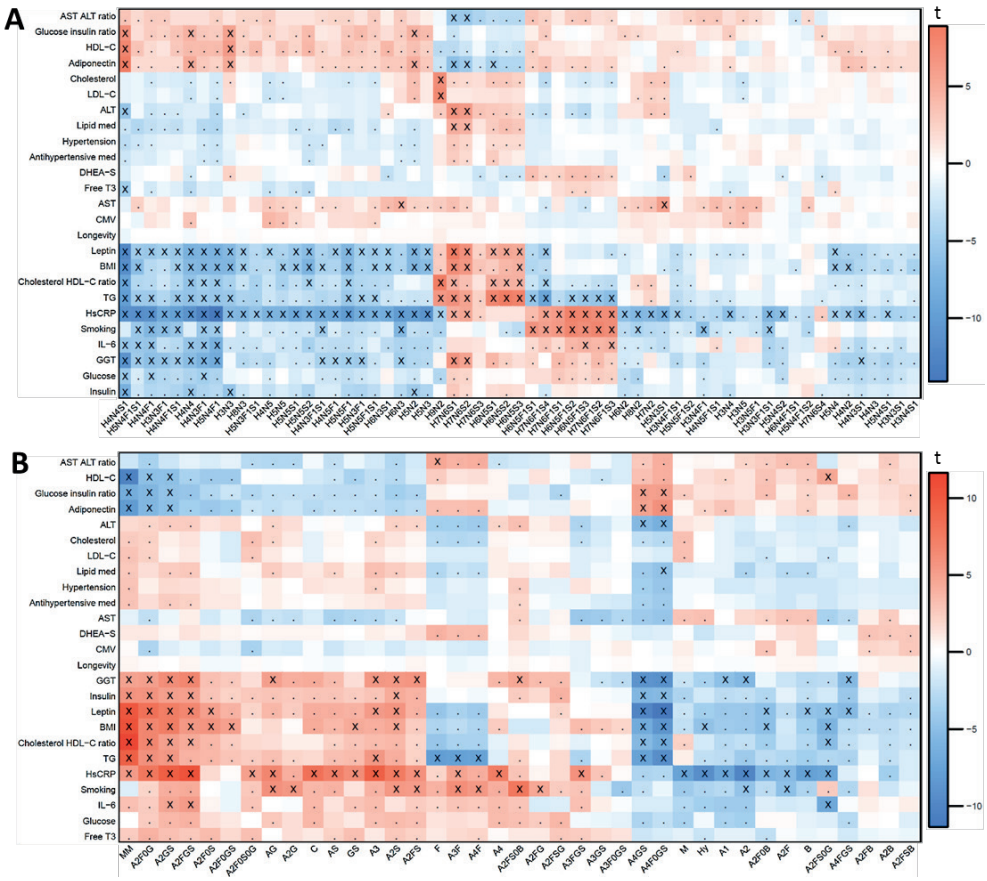

***Figure 3**. Heatmaps showing the t- (or Wald) statistics (ß/S.E.) of the associations between clinical markers of metabolic health and inflammation, and glycosylation. **A)** Single total plasma N-glycans after total area normalization. **B)** Derived glycosylation traits. All models were adjusted for age, sex, the interaction thereof, and within-family dependence. Crosses (x) indicate a statistical significance of $p <= 1.0 \times 10^{-5}$ thereby surpassing the study-wide significance threshold, whereas periods (.) indicate associations with a significance of $p <= 0.05$.*

BMI and leptin proved highly similar with respect to total plasma *N*-glycosylation associations, and showed considerable overlap with the aforementioned hsCRP as well. Taking an increase in BMI as example, galactosylation was decreased for fucosylated

diantennary species (β_{A2FS0G} = −0.55 ± 0.10; p_{A2FS0G} = 9.6 × 10^{-9}) and increased for nonfucosylated variants (β_{A2F0G} = 0.58 ± 0.09; p_{A2F0G} = 9.9 × 10^{-12}). Changes were also seen with average high-mannose size (β_{MM} = 0.89 ± 0.08; p_{MM} < 2.2 × 10^{-16}), bisection of nonfucosylated diantennary species (β_{A2F0B} = −0.39 ± 0.08; p_{A2F0B} = 2.0 × 10^{-6}), and sialylation of diantennary glycans (β_{A2GS} = 0.60 ± 0.08; p_{A2GS} = 3.1 × 10^{-13}). All of these effects were replicable within both leptin and hsCRP. However, shared with leptin but not seen with hsCRP, was the negative association of BMI with the sialylation of tetraantennary species, and specifically the nonfucosylated variants thereof (β_{A4F0GS} = −0.74 ± 0.08; p_{A4F0GS} < 2.2 × 10^{-16}). When including hsCRP and leptin as variables in the model between BMI and glycosylation, most associations between the latter two are lost with the exception of high-mannose size (β_{MM} = 0.32 ± 0.07; p_{MM} = 4.0 × 10^{-7}) and a strong remaining trend with glycan composition H4N4S1 (β_{H4N4S1} = −0.24 ± 0.07; p_{H4N4S1} = 1.8 × 10^{-4}) (**Supplemental Table S6**).

The liver marker GGT showed major associations with the total plasma *N*-glycome, whereas AST, ALT, and the ratio thereof were of only minor influence. GGT appeared similar to hsCRP in changes of increased galactosylation of nonfucosylated diantennary glycans (*e.g.*, β_{A2F0G} = 0.08 ± 0.02; p_{A2F0G} = 2.5 × 10^{-7}), sialylation of fucosylated diantennaries (*e.g.*, β_{A2FGS} = 0.09 ± 0.01; p_{A2FGS} = 9.9 × 10^{-12}), as well as an overall increased in antennarity (*e.g.*, β_{A3} = 0.07 ± 0.01; p_{A3} = 3.7 × 10^{-8}). However, the marker showed a decrease in tetraantennary sialylation similar to BMI and not hsCRP (β_{A4F0GS} = −0.11 ± 0.01; p_{A4F0GS} < 2.2 × 10^{-16}), while lacking the decreasing galactosylation of nonsialylated fucosylated diantennaries (A2FS0G) prominently seen in both BMI and hsCRP.

Clinical markers considered of beneficial metabolic nature, in our study represented by glucose/insulin, HDL-C and adiponectin[57,58], showed associations largely opposite to those established for hsCRP, BMI and GGT. Examples of this include, in case of adiponectin, the decreased galactosylation of nonfucosylated diantennaries (β_{A2F0G} = −0.06 ± 0.01; p_{A2F0G} = 3.1 × 10^{-8}), a decreased sialylation of diantennaries (β_{A2GS} = −0.05 ± 0.01; p_{A2GS} = 2.1 × 10^{-6}), a decreased average high-mannose size (β_{MM} = −0.08 ± 0.01; p_{MM} = 1.7 × 10^{-15}), and an increased sialylation of tetraantennary nonfucosylated species (β_{A4F0GS} = 0.06 ± 0.01; p_{A4F0GS} = 3.1 × 10^{-9}).

Interestingly, the only clinical markers affecting tri- and tetraantennary fucosylation (after correction for sex) proved to be smoking with a positive association (β_{A3F} = 0.50 ± 0.08; p_{A3F} = 2.6 × 10^{-10} and β_{A4F} = 0.46 ± 0.08; p_{A4F} = 2.3 × 10^{-8}), and TG levels with a negative association (β_{A3F} = −0.09 ± 0.01; p_{A3F} = 8.4 × 10^{-14} and β_{A4F} = −0.09 ± 0.01; p_{A4F} = 1.6 × 10^{-13}). In addition, smoking was also the only clinical variable to positively associate with the bisection of fucosylated nonsialylated diantennary species (β_{A2FS0B} = 0.56 ± 0.07; p_{A2FS0B} = 4.7 × 10^{-15}). Of particular interest is the high-mannose size trait (MM), which shows several of the strongest correlations, positively associating with cholesterol/HDL-C (β_{MM} = 0.08 ± 0.01; p_{MM} < 2.2 × 10^{-16}), TG levels (β_{MM} = 0.11 ± 0.01; p_{MM} < 2.2 × 10^{-16}), leptin (β_{MM} = 0.20 ± 0.02;

$p_{MM} < 2.2 \times 10^{-16}$) and BMI ($\beta_{MM} = 0.89 \pm 0.08$; $p_{MM} < 2.2 \times 10^{-16}$), whereas negatively associating with glucose/insulin ($\beta_{MM} = -0.10 \pm 0.02$; $p_{MM} = 3.0 \times 10^{-10}$), HDL ($\beta_{MM} = -0.09 \pm 0.01$; $p_{MM} < 2.2 \times 10-16$), and adiponectin ($\beta_{MM} = -0.08 \pm 0.01$; $p_{MM} = 1.7 \times 10^{-15}$). Considering the individual glycans contributing to the MM trait, the significant compositions proved to be H5N2 and H9N2.

6.5 Discussion

Here we report the MALDI-FTICR-MS analysis of the total plasma *N*-glycomes of 2144 principally healthy individuals from the LLS, and the association thereof with clinical markers of inflammation and metabolic health. In doing so, we have confirmed expectations from literature by showing a decrease of galactosylation and increase in bisection of diantennary fucosylated glycans with increasing age[59-61], and by men having a higher tri- and tetraantennary fucosylation and lower bisection than women[62,63]. Adjusting for the age and sex effects as well as the literature-reported interaction of age and sex[18,21], we proved additional associations between clinical markers for metabolic health/inflammation and plasma *N*-glycosylation characteristics including antennarity, sialylation, bisection, and galactosylation of various diantennary subgroups, and the size of high-mannose species.

Methodology

MS analysis of plasma *N*-glycosylation is not without challenges, particularly when employing MALDI ionization. A downside of this technique is the in-source and metastable loss of sialic acid residues, which is problematic given that plasma *N*-glycans are often highly sialylated (a notable exception being those from the fragment-crystallizable (Fc) region of IgG)[16,37]. Although chemical derivatization prior to MALDI-MS has been found to solve these stabilization issues (*e.g.*, by permethylation, methyl/ethyl esterification or amidation), this adds extra steps to the sample preparation workflow and often leads to byproducts[31,64-69]. Instead, here we have used MALDI-FTICR-MS equipped with an intermediate pressure source to decrease sialic acid decay[36,38]. The intermediate pressure source is known to promote ion integrity by cooling of the MALDI-generated ions[36,39-41]. In our study this has facilitated the relative quantification of *N*-glycan species containing up to four *N*-acetylneuraminic acids.

Although similarly sized investigations have analyzed plasma *N*-glycosylation by separation techniques like (U) HPLC or CGE-LIF[18-20,23,70], to our knowledge this is the largest study of its kind performed by MS[32-34]. Glycan compositions as obtained by MS provide orthogonal information to chromatographic peaks. UHPLC, for example, can separate diantennary *N*-glycan isomers with α1,3- versus α1,6-arm galactosylation and can likewise distinguish an antennary from a bisecting GlcNAc, whereas MS can provide more precise groups of di-, tri-, and tetraantennary compositions, number of sialic acids, and clear separation of high-mannose type glycans[29,71]. As such, the construction of derived glycosylation traits making

use of these features, while still biased on a compositional level, is simple to perform and provides additional insight into the complexity of glycan changes.

One example of the added information of derived traits can be found in the association of glycosylation with hsCRP. On an individual glycan level, we can only observe a relative increase in tri- and tetraantennary compositions together with the inflammation marker, and a corresponding relative decrease in all other compositions (which may be because of the total area normalization). However, when we mathematically take several individual glycans with shared biological features out of the total plasma *N*-glycome and compare them relative to each other in the form of a derived trait, we now additionally reveal, for example, a decrease in galactosylation within the subset of glycan compositions predominantly occurring on IgG-Fc (nonsialylated fucosylated diantennary species; A2FS0), a finding expected from literature[16,72,73].

Additional aspects of the analytical methodology need to be considered to allow valid interpretation of the presented findings. First, *N*-glycans are released from their protein backbones, and thus information is lost whether an observed glycan change originates from protein glycosylation or from glycoprotein abundance. Second, mass spectrometry does not distinguish isomers, meaning that a given monosaccharide mass, *e.g.*, a hexose, is assigned differently based on literature knowledge of its compositional context (*e.g.*, as mannose for H5N2 and as galactose for H4N4F1)[16,27,45,46]. Similarly, literature is the main source of information on the linkages between monosaccharides, for example presuming bisection for compositions having five *N*-acetylhexosamines but less than three galactoses (*e.g.*, H5N5F1). Although this group indeed encompasses bisection, the information within will also be confounded by triantennary structures with incomplete galactosylation, even if these are not abundant in human plasma[16,27]. Third, the methodology presented here will not provide biologically true relative ratios of glycan compositions, as the profiles will, for example, be skewed by the ionization advantage of sialylated species in negative ion mode MS. Nonetheless, the relative signal differences will still be representative for the biological directions of change, as well as providing an estimate of the magnitude.

Clinical findings

With the help of literature, we can speculate on the biological background of the observed associations. Acute inflammation, represented in our study by hsCRP and IL-6, shows to confirm previous glycomics studies with regard to the increase in antennarity, fucosylation of triantennary species, and a decrease in galactosylation of IgG glycans which is particularly well-established[9,17,21,24,59,74]. The first two may be explained by an inflammation-induced increase in plasma levels and glycosylation changes of acute phase proteins such as alpha-1-antitrypsin and alpha-1-acid glycoprotein (orosomucoid-1)[9,75]. Not only are these carriers of *N*-glycans with two or more antennae at baseline conditions, they furthermore display increased antennarity and sialyl-Lewis X upon acute inflammation[76-78]. Additional

observations include the increase of galactosylation and sialylation per galactose of diantennary fucosylated (A2F) species in general, which is notably different from the behavior of IgG. A contributor to this observation could be the level of IgM, an abundant immunoglobulin shown to increase with many autoimmune and inflammatory conditions and which carries the required highly galactosylated and sialylated A2F species[79,80].

Of particular interest is the highly significantly increased size of high-mannose glycans (MM) which is not only observed with increased hsCRP, but as well with increasing BMI, non-HDL cholesterol and TG. This glycosylation change appears to largely stem from the increase in the single glycan composition H9N2, an analyte difficult to assess by commonly used liquid chromatography with fluorescence detection[23,81]. Within the total plasma *N*-glycome, this large high-mannose glycan may predominantly originate from apolipoprotein B, the main protein constituent of most non-HDL lipoproteins (*e.g.*, VLDL, LDL)[16,82,83]. The observed correlations with H9N2 and MM can represent differing apolipoprotein B levels or glycosylation thereof, and may be indicative of an unhealthy glycosylation profile with regard to lipid transport and metabolism. The association of MM with inflammatory marker hsCRP is likely a consequence of the connection between inflammation and obesity[84], as a model corrected for BMI no longer shows significant association between high mannose size and inflammation (data not shown). The effect size being less pronounced with hsCRP than with BMI is explainable by the aforementioned increase in IgM, which contains the smaller high-mannose compositions H5N2 and H6N2 at its Asn^{279} site[80]. On the other hand, the association between MM and BMI remains true in a model adjusted for CRP and leptin, indicating the glycosylation trait may have potential to discriminate healthy from unhealthy obese.

Although largely following an inflammatory glycosylation profile, BMI, leptin, non-HDL cholesterol and TG do show a unique negative association with the sialylation of nonfucosylated tetraantennary compositions (A4F0GS), a change mainly occurring because of the relative increase of the lowly sialylated H7N6S1 and H7N6S2. Many proteins may contribute to this decrease in sialylation, a notable one being alpha-1-acid glycoprotein[85-87], which is known to bind lipophilic compounds with affinity modulated by its degree of sialylation[88-90]. This effect also applies to binding of hydrophobic drugs, possibly explaining why we similarly observed trends of decreased A4F0GS with the usage of lipid- and antihypertensive medication[89,90]. Additionally, increased levels of glycoproteins with exposed galactose residues can reflect a modulation of protein turnover, e.g. recycling by the liver-based asialoglycoprotein receptors[5,91-93]. Low sialylation would induce rapid turnover of lipid scavengers like alpha-1-acid glycoprotein, a situation which is beneficial when the blood needs to be cleared of high levels of lipophilic compounds.

Smoking proved the only phenotype to positively associate with the fucosylation of tri- and tetraantennary species (A3F and A4F), as well as with the bisection of IgG-Fc type glycans

(A2FS0B)[20,94]. Likely these reflect the chronic response to vascular injury obtained from smoking-induced shear stress and oxidative damage[95]. Increased fucosylation of acute phase proteins, when antennary-linked in the form of sialyl-Lewis X or A, facilitates recruitment to sites of injury, *e.g.*, by interactions with selectins presented on inflammation-activated endothelial cells[16,76,96].

Although it is invalid to interpret absence of statistical significance as an absence of association, we have nevertheless failed to identify derived traits or single glycans to be predictive of DHEA-S, CMV infection or the propensity for longevity. The latter is a peculiar absence, as plasma *N*-glycosylation analysis of the same cohort by HPLC had revealed two chromatographic peaks to be predictors of the phenotype[18]. Reasons for this lack of biological reproduction may include the measurement error (HPLC tends to provide more robust measurements than mass spectrometry)[71], the chromatographic peaks representing a culmination of multiple mass spectrometric compositions that are not individually significant, the previous findings being incidental, or the inability of mass spectrometry to separate or detect the responsible analytes.

Glycosylation is of high interest for the development or improvement of biomarkers for disease and patient stratification[14,97]. In particular, the findings within this study may be of benefit to the detection and discrimination of metabolic syndrome or inflammatory disorders, and may bolster the predictability of existing biomarkers such as the Framingham Risk Score[98]. Important glycosylation phenotypes in this regard would then be the derived traits MM, A3F, A4F, A2F0G, A2F0B, A2FS0G, A2FGS, and A4F0GS, as well as the various individual glycans comprising these groups, *e.g.*, most tri- and tetraantennary compositions, high-mannose compositions H5N2, H6N2, and H9N2, and the truncated *N*-glycans suggested by compositions H3N3, H4N4, and H4N4S1. Interestingly, although this last example, H4N4S1, is a glycan composition difficult to characterize in a derived trait, it does show to be the major single glycan to positively associate with most beneficial markers of metabolic health (*e.g.*, glucose/insulin, HDL-C, adiponectin) and negatively with most detrimental ones (*e.g.*, hsCRP, BMI, GGT, hypertension, TG). The protein source of the *N*-glycan is as of yet unclear, but is of interest to study in more detail.

To summarize, we have reported a large number of associations between total plasma *N*-glycosylation as measured by MALDI-FTICR-MS and clinical markers of metabolic health and inflammation. By this, we have identified glycan compositions and derived traits indicative of overall metabolic health and inflammation, as well as finding glycosylation traits uniquely associating with single marker variables. With this knowledge, we hope to contribute to the interpretation of the plasma *N*-glycome as biomarker for health and disease, and to assist clinical translation of mass spectrometric glycosylation analysis.

6.6 Supporting information

Supplemental Figures (.pdf) - Supplemental figures.

Supplemental Tables (.xlsx) - Supplemental tables.

6.7 Acknowledgments

This work was supported by the European Union Seventh Framework Programme projects HighGlycan (278535), MIMOmics (305280), and IDEAL (259679). In addition, financial support was provided by the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology and the Netherlands Consortium for Healthy Aging (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative, the Netherlands Organization for Scientific Research (NWO) and by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007).

6.8 References

1 Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130 (1993).

2 Varki, A. *et al. Essentials of Glycobiology*. 2nd edn, (2009).

3 Xu, C. C. & Ng, D. T. W. Glycosylation-directed quality control of protein folding. *Nat Rev Mol Cell Bio* **16**, 742-752, doi:10.1038/nrm4073 (2015).

4 Kontermann, R. E. Strategies for extended serum half-life of protein therapeutics. *Curr Opin Biotech* **22**, 868-876, doi:10.1016/j.copbio.2011.06.012 (2011).

5 Yang, W. H. *et al.* An intrinsic mechanism of secreted protein aging and turnover. *Proc Natl Acad Sci U S A* **112**, 13657-13662, doi:10.1073/pnas.1515464112 (2015).

6 Ferrara, C. *et al.* Unique carbohydrate-carbohydrate interactions are required for high affinity binding between Fc gamma RIII and antibodies lacking core fucose. *Proc Natl Acad Sci U S A* **108**, 12669-12674, doi:10.1073/pnas.1108455108 (2011).

7 Varki, A. & Gagneux, P. Multifarious roles of sialic acids in immunity. *Ann Ny Acad Sci* **1253**, 16-36, doi:10.1111/j.1749-6632.2012.06517.x (2012).

8 Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540-555, doi:10.1038/nrc3982 (2015).

9 Arnold, J. N., Saldova, R., Hamid, U. M. A. & Rudd, P. M. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **8**, 3284-3293, doi:10.1002/pmic.200800163 (2008).

10 Axford, J. S. Glycosylation and rheumatic disease. *Bba-Mol Basis Dis* **1455**, 219-229, doi:Doi 10.1016/S0925-4439(99)00057-5 (1999).

11 Bondt, A. *et al.* Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. *J Proteome Res* **12**, 4522-4531, doi:10.1021/pr400589m (2013).

12 Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **13**, 448-462, doi:10.1038/nrm3383 (2012).

13 Maverakis, E. *et al.* Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *J Autoimmun* **57**, 1-13, doi:10.1016/j.jaut.2014.12.002 (2015).

14 Ruhaak, L. R., Miyamoto, S. & Lebrilla, C. B. Developments in the Identification of Glycan Biomarkers for the Detection of Cancer. *Molecular & Cellular Proteomics* **12**, 846-855, doi:10.1074/mcp.R112.026799 (2013).
15 Klein, A. Human total serum N-glycome. *Adv Clin Chem* **46**, 51-85 (2008).
16 Clerc, F. *et al.* Human plasma protein N-glycosylation. *Glycoconj J* **33**, 309-343, doi:10.1007/s10719-015-9626-2 (2016).
17 Ruhaak, L. R. *et al.* Decreased Levels of Bisecting GlcNAc Glycoforms of IgG Are Associated with Human Longevity. *PLoS One* **5**, doi:ARTN e1256610.1371/journal.pone.0012566 (2010).
18 Ruhaak, L. R. *et al.* Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res* **10**, 1667-1674, doi:10.1021/pr1009959 (2011).
19 Lu, J. P. *et al.* Screening Novel Biomarkers for Metabolic Syndrome by Profiling Human Plasma N-Glycans in Chinese Han and Croatian Populations. *J Proteome Res* **10**, 4959-4969, doi:10.1021/pr2004067 (2011).
20 Knezevic, A. *et al.* Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology* **20**, 959-969, doi:10.1093/glycob/cwq051 (2010).
21 Kristic, J. *et al.* Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci* **69**, 779-789, doi:10.1093/gerona/glt190 (2014).
22 Vanhooren, V. *et al.* N-glycomic changes in serum proteins during human aging. *Rejuv Res* **10**, 521-531, doi:10.1089/rej.2007.0556 (2007).
23 Igl, W. *et al.* Glycomics meets lipidomics-associations of N-glycans with classical lipids, glycerophospholipids, and sphingolipids in three European populations. *Mol Biosyst* **7**, 1852-1862, doi:10.1039/c0mb00095g (2011).
24 Knezevic, A. *et al.* Variability, Heritability and Environmental Determinants of Human Plasma N-Glycome. *J Proteome Res* **8**, 694-701, doi:10.1021/pr800737u (2009).
25 Trbojevic Akmacic, I. *et al.* Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. *Inflamm Bowel Dis* **21**, 1237-1247, doi:10.1097/MIB.0000000000000372 (2015).
26 Novokmet, M. *et al.* Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep* **4**, 4347, doi:10.1038/srep04347 (2014).
27 Saldova, R. *et al.* Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J Proteome Res* **13**, 2314-2327, doi:10.1021/pr401092y (2014).
28 Ruhaak, L. R., Uh, H. W., Deelder, A. M., Dolhain, R. E. & Wuhrer, M. Total plasma N-glycome changes during pregnancy. *J Proteome Res* **13**, 1657-1668, doi:10.1021/pr401128j (2014).
29 Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom Rev* **18**, 349-450, doi:10.1002/(SICI)1098-2787(1999)18:6<349::AID-MAS1>3.0.CO;2-H (1999).
30 Canis, K. *et al.* Mapping the N-glycome of human von Willebrand factor. *Biochem J* **447**, 217-228, doi:10.1042/BJ20120810 (2012).
31 Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal Chem* **86**, 5784-5793, doi:10.1021/ac500335t (2014).
32 Kang, P. *et al.* 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, doi:10.1016/j.ijms.2010.11.007 (2011).
33 Borelli, V. *et al.* Plasma N-Glycome Signature of Down Syndrome. *J Proteome Res* **14**, 4232-4245, doi:10.1021/acs.jproteome.5b00356 (2015).
34 Jansen, B. C. *et al.* Pregnancy-associated serum N-glycome changes studied by high-throughput MALDI-TOF-MS. *Sci Rep* **6**, 23296, doi:10.1038/srep23296 (2016).

35 Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* **14**, 79-84, doi:10.1038/sj.ejhg.5201508 (2006).
36 Asakawa, D., Calligaris, D., Zimmerman, T. A. & De Pauw, E. In-Source Decay during Matrix-Assisted Laser Desorption/Ionization Combined with the Collisional Process in an FTICR Mass Spectrometer. *Anal Chem* **85**, 7809-7817, doi:10.1021/ac401234q (2013).
37 Powell, A. K. & Harvey, D. J. Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption ionization mass spectrometry. *Rapid Commun Mass Sp* **10**, 1027-1032 (1996).
38 Selman, M. H. J. *et al.* Immunoglobulin G Glycopeptide Profiling by Matrix-Assisted Laser Desorption Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal Chem* **82**, 1073-1081, doi:10.1021/ac9024413 (2010).
39 Lee, H., An, H. J., Lerno, L. A., German, J. B. & Lebrilla, C. B. Rapid profiling of bovine and human milk gangliosides by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry. *Int J Mass Spectrom* **305**, 138-150, doi:10.1016/j.ijms.2010.10.020 (2011).
40 Park, Y. M. & Lebrilla, C. B. Application of Fourier transform ion cyclotron resonance mass spectrometry to oligosaccharides. *Mass Spectrom Rev* **24**, 232-264, doi:10.1002/mas.20010 (2005).
41 O'Connor, P. B., Mirgorodskaya, E. & Costello, C. E. High pressure matrix-assisted laser desorption/ionization Fourier transform mass spectrometry for minimization of ganglioside fragmentation. *J Am Soc Mass Spectr* **13**, 402-407, doi:Pii S1044-0305(02)00351-3Doi 10.1016/S1044-0305(02)00351-3 (2002).
42 Westendorp, R. G. J. *et al.* Nonagenarian Siblings and Their Offspring Display Lower Risk of Mortality and Morbidity than Sporadic Nonagenarians: The Leiden Longevity Study. *J Am Geriatr Soc* **57**, 1634-1637, doi:10.1111/j.1532-5415.2009.02381.x (2009).
43 Friedewald, W. T., Levy, R. I. & Fredrickson, D. S. Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, Without Use of the Preparative Ultracentrifuge. *Clin Chem* **18**, 499 (1972).
44 Ruhaak, L. R. *et al.* Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. *Anal Chem* **80**, 6119-6126, doi:10.1021/ac800630x (2008).
45 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).
46 Freeze, H. H. Genetic defects in the human glycome. *Nat Rev Genet* **7**, 537-551, doi:10.1038/nrg1894 (2006).
47 Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918-920, doi:10.1038/nbt.2377 (2012).
48 Jansen, B. C. *et al.* MassyTools: A High-Throughput Targeted Data Processing Tool for Relative Quantitation and Quality Control Developed for Glycomic and Glycoproteomic MALDI-MS. *J Proteome Res* **14**, 5088-5098, doi:10.1021/acs.jproteome.5b00658 (2015).
49 Varki, A. *et al.* Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324, doi:10.1093/glycob/cwv091 (2015).
50 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).
51 RCore Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/. (2014).
52 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, doi:10.1093/biostatistics/kxj037 (2007).

53 Liang, K. Y. & Zeger, S. L. Longitudinal Data-Analysis Using Generalized Linear-Models. *Biometrika* **73**, 13-22, doi:DOI 10.1093/biomet/73.1.13 (1986).
54 Bladergroen, M. R. *et al.* Automation of High-Throughput Mass Spectrometry-Based Plasma N-Glycome Analysis with Linkage-Specific Sialic Acid Esterification. *J Proteome Res* **14**, 4080-4086, doi:10.1021/acs.jproteome.5b00538 (2015).
55 Heinrich, P. C., Castell, J. V. & Andus, T. Interleukin-6 and the Acute Phase Response. *Biochemical Journal* **265**, 621-636 (1990).
56 Vigushin, D. M., Pepys, M. B. & Hawkins, P. N. Metabolic and Scintigraphic Studies of Radioiodinated Human C-Reactive Protein in Health and Disease. *Journal of Clinical Investigation* **91**, 1351-1357, doi:Doi 10.1172/Jci116336 (1993).
57 O'Neill, S., Bohl, M., Gregersen, S., Hermansen, K. & O'Driscoll, L. Blood-Based Biomarkers for Metabolic Syndrome. *Trends Endocrin Met* **27**, 363-374, doi:10.1016/j.tem.2016.03.012 (2016).
58 Renaldi, O. *et al.* Hypoadiponectinemia: a risk factor for metabolic syndrome. *Acta Med Indones* **41**, 20-24 (2009).
59 Dall'Olio, F. *et al.* N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. *Ageing Res Rev* **12**, 685-698, doi:10.1016/j.arr.2012.02.002 (2013).
60 Shikata, K. *et al.* Structural changes in the oligosaccharide moiety of human IgG with aging. *Glycoconj J* **15**, 683-689, doi:Doi 10.1023/A:1006936431276 (1998).
61 Yamada, E., Tsukamoto, Y., Sasaki, R., Yagyu, K. & Takahashi, N. Structural changes of immunoglobulin G oligosaccharides with age in healthy human serum. *Glycoconj J* **14**, 401-405, doi:Doi 10.1023/A:1018582930906 (1997).
62 Ding, N. *et al.* Human serum N-glycan profiles are age and sex dependent. *Age Ageing* **40**, 568-575, doi:10.1093/ageing/afr084 (2011).
63 Ruhaak, L. R. *et al.* Targeted biomarker discovery by high throughput glycosylation profiling of human plasma alpha1-antitrypsin and immunoglobulin A. *PLoS One* **8**, e73082, doi:10.1371/journal.pone.0073082 (2013).
64 Johnson, S. B. & Brown, R. E. Simplified Derivatization for Determining Sphingolipid Fatty Acyl Composition by Gas-Chromatography Mass-Spectrometry. *J Chromatogr* **605**, 281-286, doi:Doi 10.1016/0021-9673(92)85248-R (1992).
65 Morelle, W. & Michalski, J. C. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc* **2**, 1585-1602, doi:10.1038/nprot.2007.227 (2007).
66 Wheeler, S. F., Domann, P. & Harvey, D. J. Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of alpha(2 --> 3)- and alpha(2 --> 6)-isomers. *Rapid Commun Mass Spectrom* **23**, 303-312, doi:10.1002/rcm.3867 (2009).
67 Alley, W. R. & Novotny, M. V. Glycomic Analysis of Sialic Acid Linkages in Glycans Derived from Blood Serum Glycoproteins. *J Proteome Res* **9**, 3062-3072, doi:10.1021/pr901210r (2010).
68 de Haan, N. *et al.* Linkage-specific sialic acid derivatization for MALDI-TOF-MS profiling of IgG glycopeptides. *Anal Chem* **87**, 8284-8291, doi:10.1021/acs.analchem.5b02426 (2015).
69 Reiding, K. R., Lonardi, E., Hipgrave Ederveen, A. L. & Wuhrer, M. Ethyl Esterification for MALDI-MS Analysis of Protein Glycosylation. *Methods Mol Biol* **1394**, 151-162, doi:10.1007/978-1-4939-3341-9_11 (2016).
70 Ruhaak, L. R. *et al.* Optimized workflow for preparation of APTS-labeled N-glycans allowing high-throughput analysis of human plasma glycomes using 48-channel multiplexed CGE-LIF. *J Proteome Res* **9**, 6655-6664, doi:10.1021/pr100802f (2010).
71 Huffman, J. E. *et al.* Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* **13**, 1598-1610, doi:10.1074/mcp.M113.037465 (2014).

72 Collins, E. S. *et al.* Glycosylation status of serum in inflammatory arthritis in response to anti-TNF treatment. *Rheumatology* **52**, 1572-1582, doi:10.1093/rheumatology/ket189 (2013).
73 Saldova, R., Wormald, M. R., Dwek, R. A. & Rudd, P. M. Glycosylation changes on serum glycoproteins in ovarian cancer may contribute to disease pathogenesis. *Dis Markers* **25**, 219-232 (2008).
74 Gornik, O. *et al.* Changes of serum glycans during sepsis and acute pancreatitis. *Glycobiology* **17**, 1321-1332, doi:10.1093/glycob/cwm106 (2007).
75 Peracaula, R., Sarrats, A. & Rudd, P. M. Liver proteins as sensor of human malignancies and inflammation. *Proteom Clin Appl* **4**, 426-431, doi:10.1002/prca.200900170 (2010).
76 McCarthy, C. *et al.* The Role and Importance of Glycosylation of Acute Phase Proteins with Focus on Alpha-1 Antitrypsin in Acute and Chronic Inflammatory Conditions. *J Proteome Res* **13**, 3131-3143, doi:10.1021/pr500146y (2014).
77 Higai, K., Azuma, Y., Aoki, Y. & Matsumoto, K. Altered glycosylation of alpha(1)-acid glycoprotein in patients with inflammation and diabetes mellitus. *Clinica Chimica Acta* **329**, 117-125, doi:10.1016/S0009-8981(02)00427-8 (2003).
78 Degraaf, T. W., Vanderstelt, M. E., Anbergen, M. G. & Vandijk, W. Inflammation-Induced Expression of Sialyl Lewis X-Containing Glycan Structures on Alpha-1-Acid Glycoprotein (Orosomucoid) in Human Sera. *J Exp Med* **177**, 657-666, doi:DOI 10.1084/jem.177.3.657 (1993).
79 Duarte-Rey, C., Bogdanos, D. P., Leung, P. S. C., Anaya, J. M. & Gershwin, M. E. IgM predominance in autoimmune disease: Genetics and gender. *Autoimmun Rev* **11**, A404-A412, doi:10.1016/j.autrev.2011.12.001 (2012).
80 Pabst, M. *et al.* A Microarray-Matrix-assisted Laser Desorption/Ionization-Mass Spectrometry Approach for Site-specific Protein N-glycosylation Analysis, as Demonstrated for Human Serum Immunoglobulin M (IgM). *Molecular & Cellular Proteomics* **14**, 1645-1656, doi:10.1074/mcp.O114.046748 (2015).
81 Bai, L. *et al.* Plasma High-Mannose and Complex/Hybrid N-Glycans Are Associated with Hypercholesterolemia in Humans and Rabbits. *PLoS One* **11**, doi:ARTN e014698210.1371/journal.pone.0146982 (2016).
82 Garner, B. *et al.* Characterization of human apolipoprotein B100 oligosaccharides in LDL subfractions derived from normal and hyperlipidemic plasma: deficiency of alpha-N-acetylneuraminyllactosyl-ceramide in light and small dense LDL particles. *Glycobiology* **11**, 791-802 (2001).
83 Olofsson, S. O. *et al.* Apolipoprotein-B - Structure, Biosynthesis and Role in the Lipoprotein Assembly Process. *Atherosclerosis* **68**, 1-17, doi:Doi 10.1016/0021-9150(87)90088-8 (1987).
84 Ellulu, M. S., Khaza'ai, H., Rahmat, A., Patimah, I. & Abed, Y. Obesity can predict and promote systemic inflammation in healthy adults. *Int J Cardiol* **215**, 318-324, doi:10.1016/j.ijcard.2016.04.089 (2016).
85 Zhang, S., Jiang, K., Sun, C., Lu, H. J. & Liu, Y. K. Quantitative analysis of site-specific N-glycans on sera haptoglobin beta chain in liver diseases. *Acta Bioch Bioph Sin* **45**, 1021-1029, doi:10.1093/abbs/gmt110 (2013).
86 Pompach, P. *et al.* Site-specific Glycoforms of Haptoglobin in Liver Cirrhosis and Hepatocellular Carcinoma. *Molecular & Cellular Proteomics* **12**, 1281-1293, doi:10.1074/mcp.M112.023259 (2013).
87 Dage, J. L., Ackermann, B. L. & Halsall, H. B. Site localization of sialyl Lewis(x) antigen on alpha(1)-acid glycoprotein by high performance liquid chromatography electrospray mass spectrometry. *Glycobiology* **8**, 755-760, doi:DOI 10.1093/glycob/8.8.755 (1998).
88 Fournier, T., Medjoubi-N, N. & Porquet, D. Alpha-1-acid glycoprotein. *Bba-Protein Struct M* **1482**, 157-171, doi:Doi 10.1016/S0167-4838(00)00153-9 (2000).

89 Wong, A. K. L. & Hsia, J. C. Invitro Binding of Propranolol and Progesterone to Native and Desialylated Human Orosomucoid. *Canadian Journal of Biochemistry and Cell Biology* **61**, 1114-1116 (1983).

90 Ponganis, K. V. & Stanski, D. R. Factors Affecting the Measurement of Lidocaine Protein-Binding by Equilibrium Dialysis in Human-Serum. *J Pharm Sci* **74**, 57-60, doi:DOI 10.1002/jps.2600740115 (1985).

91 Morell, A. G., Gregoriadis, G., Scheinberg, I. H., Hickman, J. & Ashwell, G. The role of sialic acid in determining the survival of glycoproteins in the circulation. *J Biol Chem* **246**, 1461-1467 (1971).

92 Ashwell, G. Citation-Classic - the Role of Surface Carbohydrates in the Hepatic Recognition and Transport of Circulating Glycoproteins. *Cc/Life Sci*, 18-18 (1986).

93 Ellies, L. G. *et al.* Sialyltransferase ST3Gal-IV operates as a dominant modifier of hemostasis by concealing asialoglycoprotein receptor ligands. *Proc Natl Acad Sci U S A* **99**, 10042-10047, doi:10.1073/pnas.142005099 (2002).

94 Vasseur, J. A., Goetz, J. A., Alley, W. R. & Novotny, M. V. Smoking and lung cancer-induced changes in N-glycosylation of blood serum proteins. *Glycobiology* **22**, 1684-1708, doi:10.1093/glycob/cws108 (2012).

95 Powell, J. T. Vascular damage from smoking: disease mechanisms at the arterial wall. *Vasc Med* **3**, 21-28, doi:10.1177/1358836X9800300105 (1998).

96 Vestweber, D. & Blanks, J. E. Mechanisms that regulate the function of the selectins and their ligands. *Physiol Rev* **79**, 181-213 (1999).

97 Krishnan, S. *et al.* Combined High-Density Lipoprotein Proteomic and Glycomic Profiles in Patients at Risk for Coronary Artery Disease. *J Proteome Res* **14**, 5109-5118, doi:10.1021/acs.jproteome.5b00730 (2015).

98 Wilson, P. W. F. *et al.* Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837-1847 (1998).