

# Glycoproteogenomics: A Frequent Gene Polymorphism Affects the Glycosylation Pattern of the Human Serum Fetuin/ $\alpha$ -2-HS-Glycoprotein

## Authors

Yu-Hsien Lin, Jing Zhu, Sander Meijer, Vojtech Franc, and Albert J. R. Heck

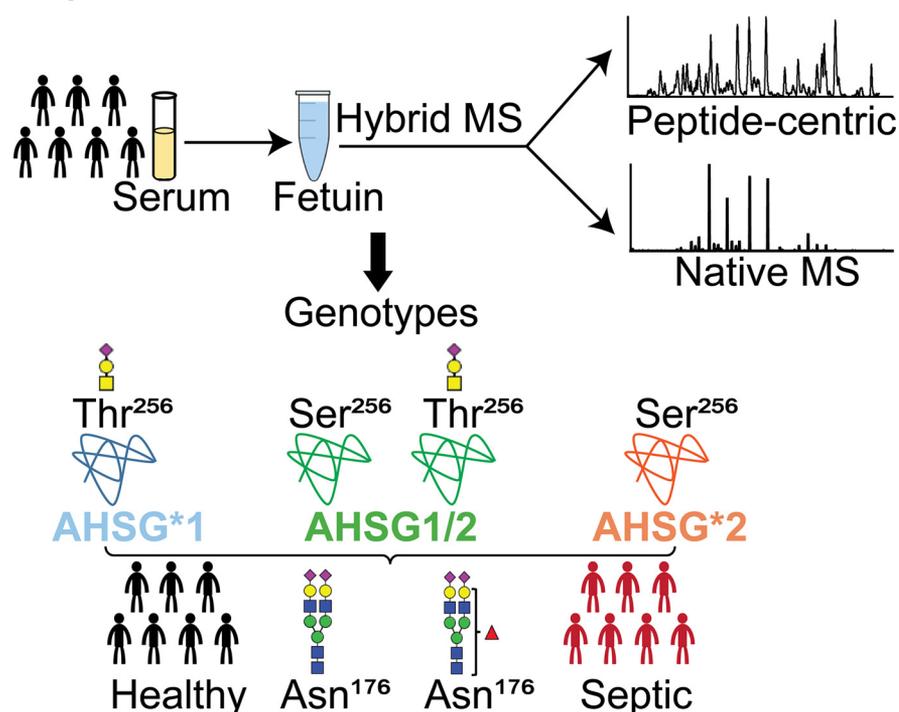
## Correspondence

v.franc@uu.nl; a.j.r.heck@uu.nl

## In Brief

Lin *et al.* applied a hybrid mass spectrometric approach to fully characterize fetuin purified from 20 individuals - 10 healthy donors and 10 septic patient donors. The results reveal that frequent genetic polymorphism influences the glycosylation pattern of fetuin. Next, the site-specific characterization of post-translational modifications on all purified fetuins allowed to classify individuals by genotype and also disease state (i.e. sepsis). These findings provide an interesting example of glycoproteogenomics, e.g. how gene polymorphism may affect glycosylation profiles.

## Graphical Abstract



## Highlights

- Frequent genetic polymorphism affects the glycosylation pattern of fetuin.
- Personalized in-depth proteoform profiling of fetuin purified from 20 donors.
- Classification of serum donors into three different genotypes.
- Septic patients show increased level of fucosylation at N-glycolation site N176.



# Glycoproteogenomics: A Frequent Gene Polymorphism Affects the Glycosylation Pattern of the Human Serum Fetuin/ $\alpha$ -2-HS-Glycoprotein<sup>§</sup>

Yu-Hsien Lin<sup>‡§</sup>, Jing Zhu<sup>‡§</sup>, Sander Meijer<sup>‡¶</sup>, Vojtech Franc<sup>‡§||</sup>,  
and  Albert J. R. Heck<sup>‡§††</sup>

Fetuin, also known as  $\alpha$ -2-HS-glycoprotein (gene name: AHS<sup>§</sup>), is one of the more abundant glycoproteins secreted into the bloodstream. There are two frequently occurring alleles of human AHS<sup>§</sup>, resulting in three genotypes (AHS<sup>§</sup>\*1, AHS<sup>§</sup>\*2, and heterozygous AHS<sup>§</sup>1/2). The backbone amino acid sequences of fetuin coded by the AHS<sup>§</sup>\*1 and AHS<sup>§</sup>\*2 genes differ in two amino acids including one known O-glycosylation site (aa position 256). Although fetuin levels have been extensively studied, the originating genotype is often ignored in such analysis. As fetuin has been suggested repeatedly as a potential biomarker for several disorders, the question whether the gene polymorphism affects the fetuin profile is of great interest. In this work, we describe detailed proteoform profiles of fetuin, isolated from serum of 10 healthy and 10 septic patient individuals and investigate potential glycoproteogenomics correlations, e.g. how gene polymorphisms affect glycosylation. We established an efficient method for fetuin purification from individuals' serum using ion-exchange chromatography. Subsequently, we performed hybrid mass spectrometric approaches integrating data from native mass spectra and peptide-centric MS analysis. Our data reveal a crucial effect of the gene polymorphism on the glycosylation pattern of fetuin. Moreover, we clearly observed increased fucosylation in the samples derived from the septic patients. Our serum proteoform analysis, targeted at one protein obtained from 20 individuals, exposes the wide variability in proteoform profiles, which should be taken into consideration when using fetuin as biomarker. Importantly, focusing on a single or few proteins, the quantitative proteoform profiles can provide, as shown here, already ample data to classify individuals by genotype and disease state. *Molecular & Cellular Proteomics* 18: 1479–1490, 2019. DOI: 10.1074/mcp.RA119.001411.

Human fetuin, also known as  $\alpha$ -2-HS glycoprotein, is an abundant glycoprotein circulating in human plasma. In adults, fetuin is secreted almost exclusively by the liver (1), but many other cells can produce this protein (2, 3). Since its discovery in 1944, human fetuin has been ascribed to various biological functions, but a concrete understanding of its exact role remains poor. One widely accepted function is its role in calcium and phosphate metabolism. Fetuin acts as one of the essential inhibitors in the prevention of ectopic mineral precipitation and deposition (4–8). Additionally, several studies have suggested a correlation between circulating fetuin levels and inflammatory, and chronic diseases, such as endotoxemia and sepsis (9), cardiovascular disease (CVD)<sup>1</sup>, type 2 diabetes (T2DM) (10, 11), and chronic kidney disease (CKD) (12), raising interest in its use as a protein biomarker.

Values of serum or plasma fetuin in healthy humans range from 300 to 750  $\mu$ g/ml (13, 14). Fetuin concentrations seem to be independent of gender, but serum fetuin levels correlate with age and might also be influenced by genetic background (15, 16). A highly frequent genetic polymorphism of the fetuin gene (AHS<sup>§</sup>), resulting from the two common alleles AHS<sup>§</sup>\*1 and AHS<sup>§</sup>\*2, was already described three decades ago (17). Genomic analysis has shown that the AHS<sup>§</sup>\*1 allele differs from AHS<sup>§</sup>\*2 by two amino acids at positions 248 and 256, respectively. The AHS<sup>§</sup>\*1 allele is characterized by ACG (Thr248) and ACC (Thr256), whereas the AHS<sup>§</sup>\*2 allele is characterized by ATG (Met248) and AGC (Ser256) (18). These two abundant alleles of AHS<sup>§</sup> result in three common genotypes (AHS<sup>§</sup>\*1, AHS<sup>§</sup>\*2, and heterozygous AHS<sup>§</sup>1/2). The population distribution of these three genotypes varies globally. The available data from several surveys demonstrate that AHS<sup>§</sup>\*1 is in frequency the foremost fetuin genotype (19–21). For example, the distribution of AHS<sup>§</sup>\*1, AHS<sup>§</sup>1/2 and

From the <sup>‡</sup>Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands; <sup>§</sup>Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands; <sup>¶</sup>Department of Molecular and Cellular Hemostasis, Sanquin Research, Amsterdam 1066 CX, the Netherlands

Received February 22, 2019, and in revised form, May 13, 2019

<sup>✂</sup> Author's Choice—Final version open access under the terms of the Creative Commons CC-BY license.

Published, MCP Papers in Press, May 16, 2019, DOI 10.1074/mcp.RA119.001411

AHSG\*2 in 697 unrelated Italians was found to be around 0.56:0.36:0.08 (22). Some studies have attempted to investigate a potential correlation between this fetuin gene polymorphism and certain diseases, however, the reported data have not been very conclusive (10, 16).

Human fetuin is a glycoprotein and its primary structure and glycosylation profile have been described (23, 24). However, the primary structure of fetuin is somehow elusive and reported discrepancies lead to confusion especially related to the absence/presence of propeptide (connecting peptide) in the matured human fetuin. According to our recent work (25), the primary structure of human fetuin contains the connecting peptide and is different from well-studied bovine fetuin. The matured human fetuin comprises an A-chain with the connecting peptide, connected to a smaller B-chain *via* a single disulfide linkage. Human fetuin is modified with a variety of posttranslational modifications (PTMs); it harbors several N- and O-glycosylations and a few phosphorylations, generating a complex mixture of proteoforms. In more detail, human fetuin harbors two N-glycosylation sites and two O-glycosylation sites in the A-chain, and one O-glycosylation site in the B-chain (25, 26). Interestingly, one A-chain O-glycan is located at position Thr256 in the AHSG\*1 genotype, which is exactly the site involved in the most frequent gene polymorphism. Whether the replacement of the Thr by the Ser at position 256 in AHSG\*2 affects the glycosylation pattern of fetuin has yet not been addressed. In this case, the Thr to Ser exchange still allows either amino residue to be O-glycosylated. In general, PTMs can modulate protein structure and play a key role in regulating the stability and physical-chemical properties of proteins (27, 28) and therefore there is a need to study proteoform profiles in detail. Human fetuin has been considered as a biomarker for certain metabolic diseases, but its quantitative analysis by various commercial enzyme-linked immunosorbent assays (ELISA) has been found to be problematic (29), which we hypothesize may be partly related to differences in the glycosylation patterns of AHSG\*1 and AHSG\*2.

Because fetuin is also known to be a negative acute phase reactant, the relationship between fetuin levels in plasma and the immune response is also of considerable interest. Glyco-

sylation patterns of acute phase proteins may change in response to diseases such as cancer (30) and during acute inflammation. Consequently, aberrant glycosylation patterns have been related to the development or progression of disease (31–33). For example, the elevated expression of sialyl-lewis X antigen has been observed on haptoglobin,  $\alpha$ -1-antichymotrypsin, and  $\alpha$ -1-acid glycoprotein in response to sepsis (34, 35). Although site-specific N-glycan profiles of these acute phase proteins have been obtained before, there has been, as far as we know, no study dealing with glycosylation changes possibly occurring in fetuin during acute and chronic inflammation.

Here, we aim to describe proteoform profiles of fetuin directly isolated from serum of healthy and septic patient individuals and fill in these knowledge gaps, thereby especially focusing on glycoproteogenomics correlations, *e.g.* how gene polymorphisms may affect glycosylation profiles. This represents a challenging task, not only because of the complicated proteoform profile of fetuin, but also because fetuin has to be purified from human serum, which contains several more abundant (glyco)proteins. We demonstrated recently that using state-of-the-art hybrid mass spectrometry (MS) approaches, combining high-resolution native MS and peptide-centric MS, we may comprehensively characterize all PTMs on various therapeutic and serum glycoproteins (36–38). Making use of such methods, we previously also reported on the very distinct proteoform profiles of fetuin obtained from three different biological sources, namely plasma-derived human fetuin from a large pool of donors, recombinant human expressed in HEK-293 cells and bovine fetuin (25). Here, we first set out to extend that latter study, focusing on fetuin extracted from individuals. We therefore first purified endogenous fetuin from several individuals' serum (healthy and septic patient donors) using a one-step mixed-mode ion-exchange chromatography fractionation (IEX) (39) and performed subsequently an in-depth characterization of the proteoform profiles. We introduce an algorithm enabling semi-automated assignment of the native MS spectra based on the data gathered from the glycopeptide-centric measurements. Our data reveal, for the first time, a dramatic effect of the gene polymorphism on the glycosylation pattern of fetuin, which needs to be considered when fetuin would be used as a biomarker. Based on this knowledge we can obtain and annotate the full proteoform profiles of fetuin from 20 individuals. These proteoform profiles allow classification into the three different genotypes. Concomitantly, looking at fetuin purified from 10 healthy individual donors and 10 septic patient donors we observe significantly increased fucosylation, especially at site N176, in the samples derived from the septic patients. Our serum fetuin analysis clearly exposes the wide variability in individual proteoform profiles. However, we demonstrate that these fetuin proteoform profiles may provide data that can be used to classify individuals by genotype and disease state.

<sup>1</sup> The abbreviations used are: CVD, cardiovascular disease; ABC, ammonium bicarbonate; AMAC, ammonium acetate; ACN, Acetonitrile; Asn, Asparagine; CAT, weak cation exchange; CKD, chronic kidney disease; dHex, deoxyhexose; EMR, Extended mass range Exactive Plus Orbitrap instrument; ELISA, enzyme-linked immunosorbent assays; ESI, electrospray ionization; ETHcD, electron-transfer combined with higher-energy collision dissociation; FA, formic acid; GlcNAc, N-acetylglucosamine; HexNAc, N-acetylhexosamine; IAA, iodoacetamide; IEX, ion-exchange chromatography; INSR, insulin receptor; Met, Methionine; MS, mass spectrometry; Neu5Ac, N-acetylneuraminic acid; Pho, phosphorylation; PTMs, post-translational modifications; Ser, Serine; SNPs, single nucleotide polymorphisms; T2DM, type 2 diabetes; Thr, Threonine; XIC, extracted ion chromatogram; WAX, weak anion exchange.

## EXPERIMENTAL PROCEDURES

**Chemicals**—Dithiothreitol (DTT), iodoacetamide (IAA), trifluoroacetic acid (TFA), formic acid (FA), ammonium bicarbonate (ABC), and ammonium acetate (AMAC) were purchased from Sigma-Aldrich (Steinheim, Germany). The fetuin sample used as reference standard was purified from a pool of human serum samples and as such acquired from Sigma ( $\alpha$ -2-HS glycoprotein; Uniprot Code: P02765, Sigma-Aldrich, Steinheim, Germany). Acetonitrile (ACN) was purchased from Biosolve (Valkenswaard, The Netherlands). Sequencing grade trypsin was obtained from Promega (Madison, WI). Gluc-C was obtained from Roche (Indianapolis, IN).

**Individual Serum Samples**—Individual human serum from 10 anonymous healthy donors was provided by Sanquin Research, Amsterdam, The Netherlands. Plasma samples were obtained after informed consent and in accordance to the ethics board of Sanquin. The whole blood from each individual was collected in a 9-ml Vacuette tube (Greiner Bio-One, Kremsmünster, Austria) containing Z Serum Clot Activator. The whole blood was then left undisturbed at room temperature for 30–60 min. The clotted material was removed by centrifugation at  $1800 \times g$  for 20 min at room temperature and the sera were transferred as 1 ml aliquots to clean 1.5 ml Eppendorf tubes, snap froze in liquid nitrogen and stored at  $-80^\circ\text{C}$ . Human serum samples from 10 septic patients were acquired from Discovery Life Science, Columbus, OH, USA. All serum samples were stored at  $-80^\circ\text{C}$  until analysis.

**Experimental Design and Statistical Rationale**—We purified fetuin and conducted native MS and peptide-centric MS analysis of serum fetuin obtained from 20 individuals (10 healthy donors and 10 septic patient donors). Our mass spectrometric data suggested that each fetuin genotype contains unique features. These features, present in all 20 serum samples, were used for classification (using a Pearson correlation coefficient) of the donors into their genotype. Following genotypic classification, we looked at the differences in fucosylation between healthy and septic donors. For that, we applied a standard *t*-test to determine if there is a statistically significant difference between these two groups of donors.

**Dual-column Ion Exchange Chromatography for Fetuin Purification from Individual Serum Samples**—An Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) consisting of a vacuum degasser, binary pump, refrigerated autosampler with 500- $\mu\text{l}$  injector loop, two column compartment with thermostat, auto collection fraction module and multi-wavelength detector, was used in this study. The dual-column set-up, comprised of a tandem WAX-CAT (PolyWAX LP,  $200 \times 2.1$  mm i.d., 5  $\mu\text{m}$ , 1000 Å; PolyCAT A,  $50 \times 2.1$  mm i.d., 5  $\mu\text{m}$ , 1000 Å) two-stage column set-up. All columns were obtained from PolyLC Inc. (Columbia, USA) (39). The column compartment was cooled to  $17^\circ\text{C}$  while the other compartments were chilled to  $4^\circ\text{C}$  to minimize sample degradation. Mobile phase Buffer A consisted of 100 mM AMAC in water, and mobile phase Buffer B consisted of 2.5 M AMAC in water. Typically, 50  $\mu\text{l}$  of serum sample was injected per run ( $\sim 3.5$  mg total protein). Elution was achieved using a multi-step gradient, consisting of six transitions with increasing proportions of Buffer B: (step 1; equilibration) 0%B, 0–6 min; (step 2; salt gradient) 0–20%B, 6–11 min; (step 3; salt gradient) 20–36%B, 11–24 min; (step 4; high salt rinse) 36–100%B, 24–28 min; (step 5; high salt wash) 100%B, 28–32 min; (step 6; restoration) 100–0%B. The flow rate was set to 800  $\mu\text{l}/\text{min}$ . The chromatograms were monitored by absorption at 280 nm and time-based fractions collected every 0.5 min from 19–23 min using an automated fraction collector.

**Native MS Analysis of Fetuin**—The human fetuin fraction, well separated from the most abundant serum proteins IgG and albumin, was collected from in between 22.5–23 min (see supplemental Fig. S1). This fraction, which contained about 25–30  $\mu\text{g}$  of the fetuin protein, was buffer exchanged into 150 mM AMAC (pH 7.2) by ultra-

filtration (vivaspin500, Sartorius Stedim Biotech, Germany) with a 10 kDa cut-off filter. The protein concentration was measured by absorbance at 280 nm and adjusted to 2–3  $\mu\text{M}$  prior to native MS analysis.

Collected fetuin fraction was analyzed on a modified Exactive Plus Orbitrap instrument with extended mass range (EMR) (Thermo Fisher Scientific, Bremen) using a standard *m/z* range of 500–10,000, as described in detail previously (40, 41). The voltage offsets on the transport multipoles and ion lenses were manually tuned to achieve optimal transmission of protein ions at elevated *m/z*. Nitrogen was used in the HCD cell at a gas pressure of  $6\text{--}8 \times 10^{-10}$  bar. The MS parameters were used typically: spray voltage 1.2–1.3 V, source fragmentation 30 V, source temperature  $250^\circ\text{C}$ , collision energy 30 V, and resolution (at *m/z* 200) 17,500. The mass spectrometer was calibrated using Csl clusters as described previously (40).

The accurate masses of observed proteoforms of fetuins were extracted by deconvoluting the native mass spectrum to zero-charge using the Intact Mass software (Protein Metrics ver. 1.5) (42). For the analysis of the PTMs, the data was processed manually, and glycan structures were retrieved based on known biosynthetic pathways. The average masses were used for these calculations; hexose/mannose/galactose (Hex/Man/Gal, 162.1424 Da), N-acetylhexosamine/N-acetylglucosamine (HexNAc/GlcNAc, 203.1950 Da), deoxyhexose (dHex, 146.1430 Da), N-acetylneuraminic acid (Neu5Ac, 291.2579 Da) and phosphorylation (Pho 79.9799 Da). All used symbols and nomenclature are based on the recommendations of the Consortium for Functional Glycomics (43).

**In-solution Digestion for Peptide-centric Proteomics**—Around 2  $\mu\text{g}$  of fetuin protein was taken from the collected fraction and dissolved into 50 mM ABC at a concentration of 1 mg/ml. The sample was then reduced with 4 mM DTT at  $56^\circ\text{C}$  for 30 min and alkylated with 8 mM IAA at room temperature for 30 min in the dark. Fetuin was digested for 3 h with Glu-C at an enzyme-to-protein-ratio of 1:75 (w/w) at  $37^\circ\text{C}$  and the resulted peptide mixtures were further digested by using trypsin (1:100; w/w). All proteolytic digests containing modified glycopeptides were desalted using Oasis HLB C18 cartridges, dried, and reconstituted in 20  $\mu\text{l}$  of 0.1% FA before liquid chromatography (LC)-MS and MS/MS analysis.

**LC-MS and MS/MS Analysis**—All peptides (300 fmol of fetuin peptides) were separated and analyzed using an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) coupled on-line to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Reversed-phase separation was accomplished using a 100  $\mu\text{m}$  inner diameter 2 cm trap column (in-house packed with ReproSil-Pur C18-AQ, 3  $\mu\text{m}$ ) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50  $\mu\text{m}$  inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7  $\mu\text{m}$ ) (Agilent Technologies, Amstelveen, The Netherlands). Mobile-phase solvent A consisted of 0.1% formic acid in water, and mobile-phase solvent B consisted of 0.1% formic acid in ACN. The flow rate was set to 300 nL/min. A 45 min gradient was used as follows: 0–5 min, 100% solvent A; 13–44% solvent B within 20 min; 44–100% solvent B within 3 min; 100% solvent B for 1 min; 100% solvent A for 17 min. For the MS scan, the mass range was set from 375–1500 *m/z* at a resolution of 120,000 and the AGC target was to  $4 \times 10^5$ . For the MS/MS measurements, electron-transfer combined with higher-energy collision dissociation (ET<sub>h</sub>CD) (44) was used and performed with a normalized collision energy of 35%. For the MS/MS scan, the mass range was set from 125–2000 *m/z*; the AGC target was set to  $5 \times 10^4$ . The precursor isolation width was 1.6 Da, and the maximum injection time was set to 200 ms.

**LC-MS and MS/MS Data Analysis**—The raw data files containing MS/MS spectra of fetuin peptides were processed using Byonic software (ver 3.2.0) (Protein Metrics Inc.) (45). The following parameters were used for data searches in Byonic: precursor ion mass

tolerance, 10 ppm; product ion mass tolerance, 20 ppm; fixed modification, Cys carbamidomethyl; variable modification: Met oxidation, Ser, Thr and Tyr phosphorylation, and both N- and O-glycosylation from mammalian glycan databases; the allowed number of miscleavages was set to 4. Trypsin (C-terminal RK) and Glu-C (C-terminal DE) enzyme specificity search was chosen for all samples. The fasta file used for the peptide searches contained both the AHSG\*1 and AHSG\*2 fetuin amino acid sequences (UniProtKB - P02765, FETUA\_HUMAN, AHSG\*1 has Thr-248/Thr-256; AHSG\*2 has Met-248/Ser-256). Byonic peptide cut-off score of 200 was used and all PTM-modified identified spectra were further manually inspected. Site-specific quantification of the fetuin PTMs was performed as follows; the first three isotopes were taken from each manually validated peptide proteoform for the calculation of the peak areas. Each peptide that contained modified sites was normalized individually so that the sum of all its proteoform areas was set to 100%. The average peptide ratios from all measurements were taken as a final estimation of the abundance. The XICs were obtained using the software Skyline (ver 4.2.0.18305) (46). The glycan structures of each glycoform were manually annotated. Hereby, reported glycan structures are depicted without the linkage type of the glycan units because our acquired MS/MS data do not directly provide such information.

**Automated Site-specific Proteoform Annotation in Native MS Spectra Using Peptide-centric Data**—We annotated in a site-specific manner each proteoform represented by peaks in the native MS spectra using the relative abundance of all PTMs sites derived from the peptide-centric data (the algorithm is made publicly available and supplemented at <https://github.com/juer120/NativePTMannotation>). The process of annotation involved two major steps: (1) the relative abundances of all possible PTM combinations on the proteoforms are retrieved from the peptide-centric data; (2) the masses of all proteoforms are extracted from the native MS spectra and then matched with the calculated proteoforms consisting of the PTM combination with the highest relative abundances, and mass window centered around detected proteoform within 100 ppm.

In detail, we first input the mass of the protein backbone retrieved from the protein sequence corrected for the presence of disulfide bridges ( $F_{pb}$ ) and then calculate the theoretical  $m/z$  of all proteoforms in determined charge state ( $z$ ) by permuting masses of all PTMs identified from peptide-centric measurements. The  $m/z$  of a proteoform  $F^{z+}$  can be calculated as:

$$F^{z+} = (F_{pb} + \sum_{i=1}^n f_{ij} + z \times 1,007276)/z$$

where  $n$  is the number of PTM sites;  $f_{ij}$  is the mass of the  $j$ -th modification at site  $i$ . Second, we calculate the relative abundance  $P$  of a proteoform  $F^{z+}$  with all possible PTM combinations, which is calculated by:

$$P = \prod_{i=1}^n P_{ij}$$

where  $P_{ij}$  is the normalized relative abundance of the  $j$ -th modification at site  $i$ ;  $n$  is the number of PTM sites.  $P_{ij}$  is calculated by:

$$P_{ij} = \frac{V_{ij}}{\sum_{j=1}^k V_{ij}}$$

where  $k$  is the number of possible PTM isoforms at site  $i$  (including unoccupied site);  $V_{ij}$  is the abundance of  $j$ -th modification at site  $i$ , which is calculated based on the XIC. Third, we input all  $m/z$  of proteoforms ( $M^{z+}$ ) extracted from raw native MS spectrum and then match them with theoretical  $m/z$  ( $F^{z+}$ ) within 100 ppm, producing a list

TABLE I  
Overview of donor characteristics. The first column lists the used sample names

Sample name	Sex	Age	Gene type	Sample type
M1	M	48	AHSG*1	Healthy
F1	F	37	AHSG*1	Healthy
F2	F	26	AHSG*1	Healthy
F5	F	63	AHSG*1	Healthy
M2	M	55	AHSG*2	Healthy
M3	M	54	AHSG*2	Healthy
F3	F	33	AHSG1/2	Healthy
F4	F	27	AHSG1/2	Healthy
M4	M	27	AHSG1/2	Healthy
M5	M	63	AHSG1/2	Healthy
F6	F	68	AHSG*1	Sepsis
F7	F	19	AHSG*1	Sepsis
F8	F	73	AHSG*1	Sepsis
M7	M	18	AHSG*1	Sepsis
F9	F	33	AHSG*1	Sepsis
F10	F	82	AHSG*1	Sepsis
M8	M	47	AHSG1/2	Sepsis
F11	F	55	AHSG1/2	Sepsis
F12	F	33	AHSG1/2	Sepsis
F13	F	37	AHSG1/2	Sepsis

of all matched  $m/z$ . Lastly, we export the highest relative abundance of PTMs combination corresponding to the matched  $m/z$  in the list.

## RESULTS

**Native Mass Spectra and Peptide-centric Analysis of Fetuin Purified from Healthy Individuals**—We initiated our study characterizing in detail human serum fetuin proteoform profiles of 10 randomly selected healthy individuals (Table I). We soon found out that these individuals harbor different fetuin genotypes and started to investigate the effect of this genetic polymorphism on the fetuin glycosylation pattern. Because fetuin is produced by at least two frequently occurring variants of the gene AHSG, AHSG\*1, and AHSG\*2, one could expect to observe distinct proteoform profiles of fetuin isolated from these individuals. To investigate this, we first loaded 50  $\mu$ l of human serum of each individual on the dual mixed-mode column ion exchange chromatography system, comprising a weak cation exchange (CAT) column preceded by a weak anion exchange (WAX) column. Using this chromatographic set-up for serum, we collected in each run a unique fraction containing fetuin, which was subsequently concentrated, buffer exchanged, and subjected to high-resolution native MS (supplemental Fig. S2). Fig. 1 displays three representative zero-charge deconvoluted native mass spectra originating from individuals with the three different genotypes of fetuin (donor F1 - AHSG\*1, donor M3 - AHSG\*2, and donor F3 - AHSG\*1/2). The grouping of all acquired spectra was done based on signature peaks specific for each genotype, as explained in more detail below. Although all three native mass spectra appeared to be quite distinct, many of the peaks were present in each spectrum and differed from each other pri-

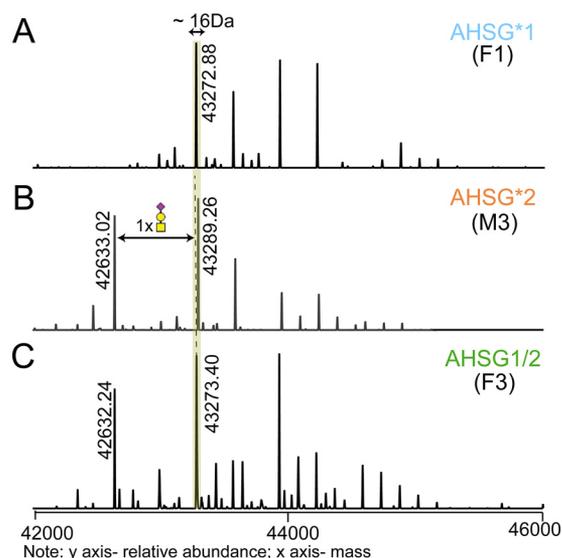


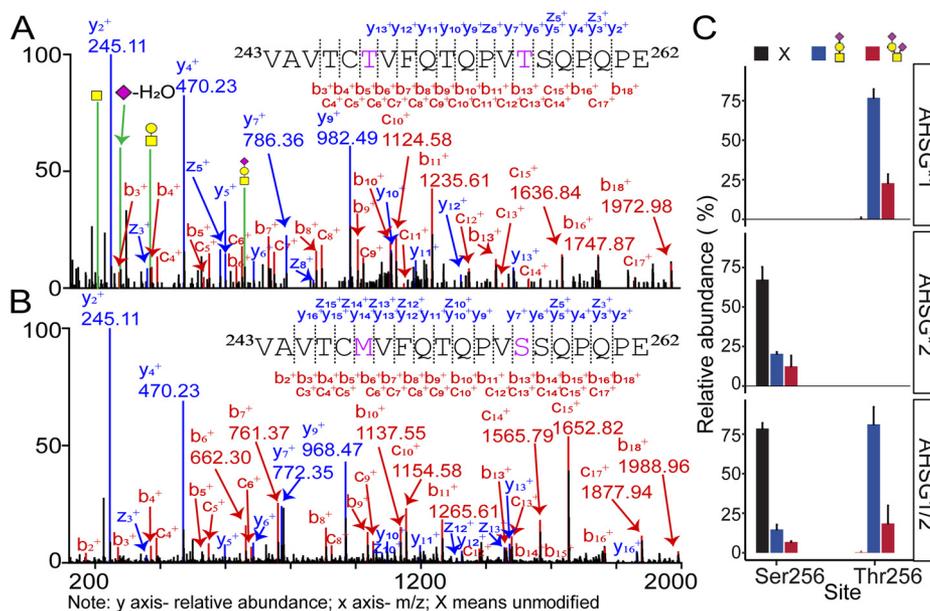
FIG. 1. Different genotypes lead to different proteoform profiles.

Three representative zero-charge deconvoluted native mass spectra of fetuin purified from serum of A, donor F1 (AHSG\*1), B, donor M3 (AHSG\*2), and C, donor F3 (AHSG1/2). Corresponding proteoforms between AHSG\*1 and AHSG\*2 differ in mass from each other by 16 Da (as shown for the proteoforms with the mass of 43,272.88 Da in A and 43,289.26 Da in B). In B, the proteoform with the mass of 42,633.02 Da differs from the most abundant proteoform by 656 Da, which agrees to an O-glycan composition of HexNAc<sub>1</sub>Hex<sub>1</sub>Neu5Ac<sub>1</sub>. This proteoform (42,632.24 Da) can be observed in C, but is fully absent in A.

marily in signal intensity. Nevertheless, a closer look into the spectra revealed some key differences. One intense peak with the mass of 42,633.02 Da detected in AHSG\*2 and AHSG\*1/2 (42,632.24 Da) was completely missing in all AHSG\*1 (Fig. 1A, and supplemental Fig. S2A). Therefore, this peak was specific for AHSG\*2 and AHSG\*1/2, and its complete absence defined unambiguously AHSG\*1. Interestingly, the missing proteoform in AHSG\*1 seemed to represent a significant difference in the proteoform profile between AHSG\*1 and AHSG\*2 and hinted at a substantial influence of the genetic polymorphism on the glycosylation profile. Additionally, when comparing corresponding peaks between AHSG\*1 and AHSG\*2, the mass difference of 16 Da could be observed (Fig. 1A and 1B, supplemental Fig. S2A and S2B). This is because of the mass difference between the backbone amino acid sequences of AHSG\*1 and AHSG\*2, exactly 15.99 Da. The native mass spectra of fetuin from the heterozygote genotype AHSG\*1/2 donors showed more complex proteoform profiles (Fig. 1C, supplemental Fig. S2C) compared with the other two genotypes. This is expected, as it should be a combination of all the proteoform profiles observed for AHSG\*1 and AHSG\*2. In theory, we should detect double peaks for all proteoforms that are common for AHSG\*1 and AHSG\*2. Nevertheless, because of the very small  $m/z$  difference between both AHSG genes (16 Da, *i.e.* 1.23 Th for the 13<sup>+</sup> charge state), we observed in practice broader or partially split peaks in the

native mass spectra obtained for the heterozygote donors. Resolving such a small mass difference would require  $\sim 10$  times higher mass resolving power in the mass region around 4000  $m/z$ . Our findings from the native mass analysis could be further supported and explained with the subsequent (glyco) peptide-centric analysis of all individual donors' fetuins. We digested fetuins from all three genotypes with trypsin and Gluc-C, which resulted in a set of peptides for subsequent sequencing by LC-MS/MS analysis. Data interpretation provided information about the site location, PTM type, composition and abundance of all identified peptides (for manually annotated MS/MS spectra see supplemental File S1 and S2; for Byonic searches and MS/MS spectra see Data Availability). Fig. 2A, 2B and supplemental File S1 displays annotated EThcD MS/MS spectra of two peptides with amino acid sequence <sup>243</sup>VAVTCTVFQTQPVTSSQPQPE<sup>262</sup> and <sup>243</sup>VAVTC-MVFQTQPVSQPQPE<sup>262</sup>, respectively signature peptides derived from proteolytic digestion of AHSG\*1 and AHSG\*2. These peptides contain the two aforementioned genotypic sites. Notably, manual inspection of the XICs obtained from all three fetuin genotypes revealed that Thr256 on the AHSG\*1 peptide was fully occupied with O-glycans containing 0, 1 or 2 sialic acids connected to the core structure HexNAc<sub>1</sub>Hex<sub>1</sub>, whereas Ser256 on the AHSG\*2 peptide was found mostly to be non-modified (Fig. 2C, supplemental Fig. S3). Thus, although Ser256 was marginally modified by O-glycan, it differs significantly from Thr256 which was found to be fully O-glycosylated. The relative abundance of all detected variants of the peptides containing these sites across the three different genotypes is displayed in Fig. 2C. The relative quantification of the site occupancy of Thr/Ser256, as obtained for all 10 studied healthy individuals, was processed using Skyline and provided in supplemental Table S1. From the peptide-centric data, we could conclusively explain the origin of the mass difference between the peaks 43,272.88 Da in AHSG\*1 and 42,633.02 Da in AHSG\*2 (HexNAc<sub>1</sub>Hex<sub>1</sub>Neu5Ac<sub>1</sub> - 15.99).

**Site-specific PTMs Annotation of the Native Mass Spectra**—We introduced earlier an algorithm facilitating the integration of the data from native mass spectrometry and peptide-centric LC-MS/MS data, which allowed us to assess the integrity of the glycopeptide characterization though *in silico* construction of an intact proteoform profile from all combined peptide-centric data (36). The algorithm lists all possible combinations of proteoform masses based on observed PTMs, at the peptide-centric level, and eventually constructs simulated native-like mass spectra, which can be correlated to the experimental native mass spectra. However, in the earlier approach the resulted list containing all predicted proteoforms, with the total PTM composition, did not contain any site-specific information on the PTMs. Here, we further extended our algorithm (made publicly available) and applied it for the semi-automated annotation of PTMs to all the peaks in the native mass spectra, now also including site location. To perform this more advanced assignment, we amended our



**FIG. 2. Peptide signatures of distinct fetuin genotypes.** EThcD MS/MS spectra of the peptide A,  $^{243}\text{VAVTCTVVFQTPVTSQPQPE}^{262}$  (AHSG\*1) and B,  $^{243}\text{VAVTCMVVFQTPVSSQPQPE}^{262}$  (AHSG\*2), obtained by proteolytic digestion (Trypsin + Glu-C), harboring the two mutation sites. The EThcD spectra provide confirmation of the unique peptide sequence and position of the O-glycan, being Thr256 in AHSG\*1. Spectra were acquired for the precursor ions with three charges and  $m/z$  of 958.78, and 745.03 for AHSG\*1 and AHSG\*2, respectively as these peptides were predominantly non-modified for AHSG\*2 and fully modified for AHSG\*1. C, Quantification of the peptide signatures, containing the mutations and O-glycosylation site Thr/Ser256. In these bars, the abundance of these peptides was averaged over all genotype-specific fetuin samples purified from the serum of the 10 healthy individuals across all three genotypes, AHSG\*1 ( $n = 4$ ), AHSG\*2 ( $n = 2$ ) and AHSG1/2 ( $n = 4$ ). Relative abundances of peptide proteoforms were estimated from their corresponding extracted ion chromatograms (XICs) and normalized to 100%. The AHSG\*1 and AHSG\*2 signature peptides contain the Thr256 and Ser256 site, respectively. Both peptides can be extracted and separately quantified for the heterozygote AHSG1/2 donors. Black bar is unmodified, blue bar is modified by HexNAc<sub>1</sub>Hex<sub>1</sub> with one sialic acid and red bar is modified by HexNAc<sub>1</sub>Hex<sub>1</sub> harboring two sialic acids.

in-house scripts written in R (see Methods section for the detailed description). In short, we first generated a library of proteoforms with specific masses and probability ranks using the data (masses and relative intensities) derived from the peptide-centric analysis. All proteoform mass combinations were ranked according to the intensity of the detected PTM peptides from which they were constructed and the match between their theoretical and experimental masses. This means that proteoforms with the same masses (within 100 ppm), but different peptide combinations, were sorted based on the better mass match and intensity of the peptides (meaning that closer mass matches and higher intensities do get higher ranks). Next, we discarded all peaks with a relative intensity of less than 5% (related to the base peak) in the native mass spectra. The masses of peaks were then matched to the generated library with the highest ranked proteoforms. This allowed us to visualize the most probable proteoforms in the native mass spectra with their PTM composition in a site-specific manner. Fig. 3 shows such a visualization for fetuin which was purified from a pool of donors of all different genotypes (our reference sample). Cumulatively, based on our experimental data, we defined on fetuin one fixed modification (Ser138 phosphorylation) and five other PTM sites (A = Asn156, B = Asn176, C = Thr/Ser256, D = Thr270, and E = Ser346). The script operated with the fetuin

backbone mass corrected for the mass shift induced by the disulfide bonds ( $-12 \times 1.0079$  Da) and one phosphate ( $+79.97$ ) moiety (Ser138). Peaks in the native mass spectra were then automatically annotated based on the library containing the PTM combinations with the highest rank. The result contained a list of proteoforms with their masses, the most probable PTM combinations and site-specific information as shown in supplemental Table S2.

*Site-specific Annotation Fully Explains Differences Observed Among AHSG\*1, AHSG\*2 and AHSG\*1/2 Fetuin*—The native mass spectra obtained from genotype-representative individuals with site-specifically annotated proteoforms defining each genotype (*i.e.* F1/AHSG\*1, M3/AHSG\*2, and F3/AHSG\*1/2) are provided in supplemental Fig. S4. Some major differences observed across all individuals defining fetuin genotypes were already indicated above. However, the exact differences among them could be only clarified after site-specific characterization of all proteoforms. The most striking one represented the high-intensity peak with a mass of 42,633.02 Da, which is completely missing in fetuin from AHSG\*1 donors and its composition consisted of two biantennary N-glycans in positions A and B ( $2 \times \text{HexNAc}_4\text{Hex}_5\text{Neu5Ac}_2$ ) and one O-glycan in position D ( $\text{HexNAc}_1\text{Hex}_1\text{Neu5Ac}_2$ ). We explained the absence of this proteoform in AHSG\*1 above, being due to the complete occupation of position C (Thr256) O-glycan-site.

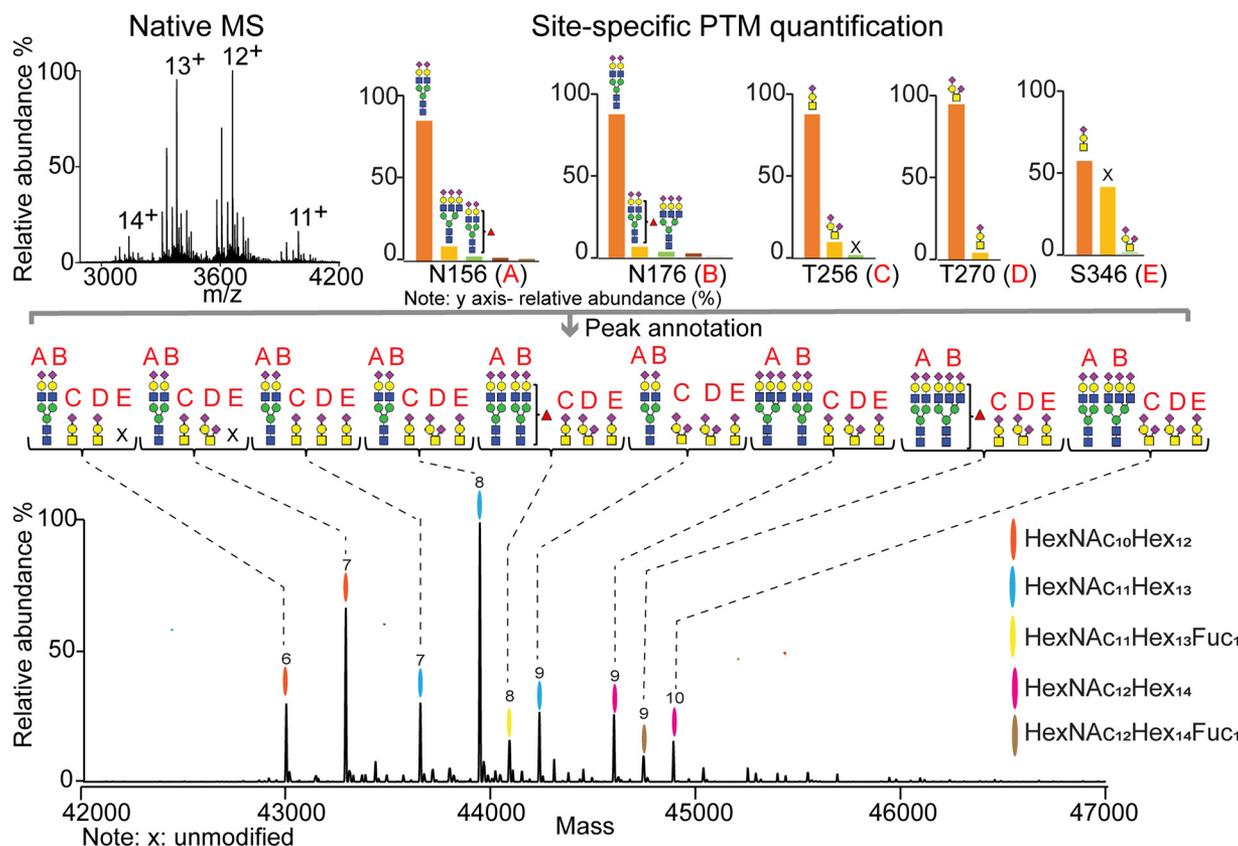
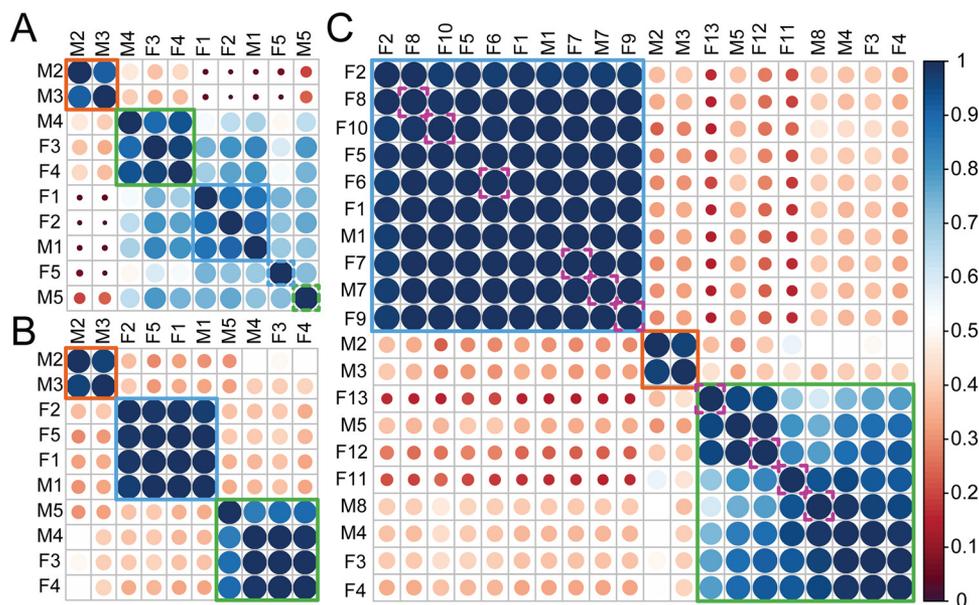


FIG. 3. **Comprehensive, quantitative and site-specific annotation of fetuin proteoform profiles.** The site-specifically annotated zero-charge deconvoluted native mass spectrum of fetuin pooled from human sera of various donors. The overall PTM compositions of the most abundant proteoforms are color coded. Each color represents a glycan composition without the sialic acids; the number of sialic acids attached is marked on the top of each peak. All displayed proteoforms contain one phosphate moiety. The site-specific annotation of the 5 glycosylation sites present on fetuin (A = Asn156, B = Asn176, C = Thr256, D = Thr270, and E = Ser346) was assigned by using our in-house developed software, making use of the integration of the native MS and peptide-centric MS data. The complete list with all annotated proteoforms and their relative abundance can be found [supplemental Table S2](#).

This was in pronounced contrast with position C (Ser256) in AHSG\*2, which was found to be occupied only partially. The second major difference between AHSG\*1 and AHSG\*2 was defined by their different protein backbone mass induced by the two mutations, resulting in the exact mass difference of 15.99 Da (37,177.01–37,193.00). This could be extracted already from the two most intensive proteoforms in the native mass spectra of F1 (AHSG\*1) and M3 (AHSG\*2) displayed in [Supplemental Fig. S4AB](#). Although both proteoforms had the same total PTM composition, site-specific annotation also revealed a difference in the O-glycan positions occupancy. Donor F1 had O-glycans at positions C and D, whereas M3 at positions D and E. A complete list of site-specifically annotated proteoforms from the three individuals representing different genotypes can be found in [supplemental Tables S3–S5](#). In F1, we could annotate 16 proteoforms ([supplemental Table S3](#)), in M3, 15 proteoforms ([supplemental Table S4](#)), and lastly in the heterozygote donor F3 34 proteoforms ([supplemental Table S5](#)). The proteoform profiles differed from each other not only by their complexity, but also by the relative abundances of peaks.

Having annotated the three representative native mass spectra, we attempted a classification of all 10 studied healthy individuals based on their proteoform profiles, following a similar approach as used earlier to classify glycol-engineered erythropoietin variants (47). We initially used unsupervised hierarchical clustering to construct a matrix based on all native mass spectra, as shown in Fig. 4A. This already resulted in three distinctive clusters, but also two clear outliers (donors F5 and M5). The first cluster represented fetuins from AHSG\*2 donors (orange box). The second (green box) and third cluster (blue box) contain fetuin from AHSG1/2 and AHSG\*1 donors, respectively. Upon closer inspection, the native mass spectra of the two outlier samples (F5 and M5), displayed more complex proteoform profiles, with nearly all peaks coexisting in pairs differing from each other by 80 Da ([supplemental Fig. S2](#)). From the LC-MS/MS data we could conclude that these two donors have a high occupancy of a second phosphorylation site, namely Ser330 ([supplemental File S2](#)). The presence of this extra phosphate moiety interfered with the genotype classification based on the complete proteoform profiles. Nevertheless, as we above identified characteristic



**FIG. 4. Classification of fetuin proteoform profiles by hierarchical clustering.** A, Unsupervised clustering of fetuin proteoform profiles derived from the 10 healthy individuals based on the correlation between their native mass spectra. B, Clustering of fetuin proteoform profiles derived from 10 healthy individuals based on specific signature peaks (see supplemental Fig. S4). Color and size of the circles represent the similarity between the proteoform profiles of different genotypes. Orange, blue and green boxes represent fetuin originating from donors representing the AHSG\*2, AHSG\*1 and AHSG1/2 genotypes, respectively. Comparing the results in A and B the classification in A is distorted by two outliers derived from donor F5 (AHSG\*1) and M5 (AHSG\*2), respectively caused by a major change in fetuin phosphorylation in these two individuals. C, Clustering of fetuin proteoform profiles derived from 10 healthy and 10 septic patients based on specific signature peaks, where pink dashed line boxes indicate fetuin derived from septic individuals.

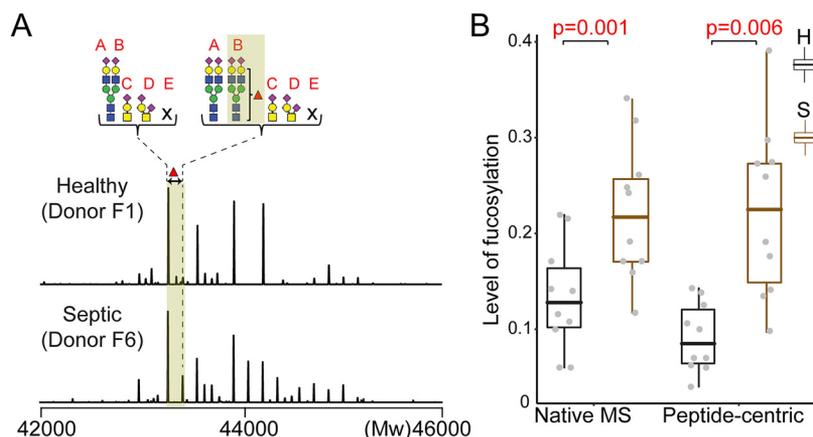
proteoforms defining each genotype, we repeated the classification, but now using only these signature peaks. The result is shown in Fig. 4B, where the orange, blue and purple boxes represent the genetic variants of fetuin. This data clearly exemplifies that, over a sample size of 10 healthy individuals, distinctive glycosylation patterns can be linked to the distinctive fetuin genotypes.

**Proteoform Profiles of Fetuin from Septic Patients Donors**— Alternation of glycosylation can be a trademark of inflammation, as reported to occur on several acute phase proteins;  $\alpha$ -1-acid glycoprotein, haptoglobin, and  $\alpha$ -1-antichymotrypsin (34, 35). However, such data for fetuin is sporadic (48). To investigate, whether fetuins' glycosylation pattern also alters upon acute inflammation, we next purified serum fetuin from 10 individual septic patient donors (Table I) and conducted the same mass spectrometric analysis as before. Our classification strategy quickly classified all 20 native mass spectra (from the 10 healthy and 10 septic patient donors) into the three genotypes (Fig. 4C). This classification could be further corroborated by the peptide-centric LC-MS/MS data (supplemental Table S6). Next, we focused on whether we could classify the healthy and septic patient donors. Fetuin obtained from septic patient donors contained some molecular species with a prominently increased intensity originating from proteoforms with fucosylated N-glycans. Zooming in to this, we focused on a specific fetuin fucosylated proteoform and its non-fucosylated variant and used the ratio of their peak in-

tensities to estimate the level of fucosylation (Fig. 5A). We calculated from the native mass spectra the average intensities of the two proteoforms using their peaks detected in the 12<sup>+</sup> and 13<sup>+</sup> charge states (supplemental Table S7). The peptide-centric LC-MS/MS analysis allowed us to pinpoint the by sepsis most affected N-glycosylation site as Asn176 (B). The ratios of peak areas of the fucosylated to nonfucosylated peptide containing this site validated the clear trend of fucosylation increase in septic patients (Fig. 5B and supplemental Table S8). This change in proteoform profile can thus be used to classify septic patients from healthy donors with good statistical values ( $p < 0.006$ ).

## DISCUSSION

Here we combined high-resolution native mass spectrometry with peptide-centric (glyco)proteomics to analyze the medium abundant serum glycoprotein fetuin, which we were able to directly purify from serum of 20 donors, 10 healthy controls and 10 septic patients. From the data, it became apparent that the samples were derived from people with distinct frequently occurring genotypes, that all result in a different set of fetuin proteoforms. We further investigated in detail the variations in proteoform profiles caused by the donors' specific genotypes, providing an interesting example of glycoproteogenomics. We next also extended our analysis to serum fetuins purified from septic patient donors.



**FIG. 5. Septic patients display enhanced fucosylation on Asn176.** *A*, Two representative zero-charge deconvoluted native mass spectra of a healthy donor (F1) and septic patient (F6). The average intensities of the site-specifically assigned fucosylated proteoform and its non-fucosylated variant were used for the determination of the relative fucosylation level. The intensities were extracted from the native mass spectra using the corresponding ion signals detected in the  $12^+$  and  $13^+$  charge states. *B*, Comparison of the extent of fucosylation obtained from the native mass spectra and peptide-centric analysis. The level of fucosylation in the peptide-centric data was determined by ratios of the peak areas of the fucosylated and non-fucosylated peptide containing the sepsis affected N-glycosylation site Asn176. Both approaches resulted in a statistically significant separation of the healthy and septic patients with  $p$  values of 0.001 and 0.006, respectively. (A = Asn156, B = Asn176, C = Thr256, D = Thr270, and E = Ser346).

*Replacement of Thr in AHSG\*1 to Ser in AHSG\*2 Changes O-glycosylation Occupancy*—In both the native MS and LC-MS/MS data we observed that the mutation in position 256, which is a Thr in AHSG\*1 and a Ser in AHSG\*2, had a dramatic effect on fetuin O-glycosylation, primarily caused by the diminished occupancy of O-glycans on AHSG\*2/Ser256. Notably, both a Thr and Ser can be modified by O-glycosylation, but our finding that the O-glycosylation is reduced on fetuin Ser256 compared with Thr256 is in line with a meta-analysis on O-glycosylation occupancy, which reported a considerable preference for Thr over Ser (49). We provide here the first experimental evidence showing how the replacement of Thr for Ser in fetuin changes the occupancy of this site. In general, changes in O-glycosylation may affect protein stability or protection against proteolytic activity (50). Impact of gene mutation, including SNPs, on protein glycosylation has been reported earlier, albeit only in a few studies (51–53). Notably, such mutations have been associated with altered physiological states, especially the ones leading to disease. Whether different the here reported proteoform profiles of fetuin have biological implications needs to be further investigated.

*Coupling Between Fetuin Backbone Processing and Phosphorylation*—The native mass spectra of two healthy donors, F5 and M5, attracted our attention because of their quite distinctive proteoform profiles, representing the two outliers in Fig. 4A. We observed that this was primarily caused by the site-occupancy of the additional fetuin phosphorylation site Ser330. Human serum fetuin is predominantly present as singly phosphorylated, with a site-occupancy of nearly 100% at Ser138 (25). Earlier phosphoproteomics analysis has revealed that at least four more Ser residues can be partially phosphorylated (54–56). In a recent study comparing obese

versus lean individuals, phosphorylation on fetuin Ser330 was attributed with a metabolically active pool of human fetuin, within the larger total pool of circulating fetuin (57). Interestingly, this Ser330 phosphorylation site is in the 40-amino acid residue long propeptide (also termed “connecting peptide”), which is believed to be lost during posttranslational processing in serum. Data from literature has suggested that fetuin can indeed exist in human serum in two processed forms, with- or without the connecting peptide (58). One putative proteolytic cleavage separating fetuin into two chains occurs at site 340 leading to the removal of an Arg residue. The second proteolytic cleavage releasing the connecting peptide (with the phosphate at Ser330) is Leu300. It is therefore tempting to speculate about a possible cross-talk mechanism between Ser330 phosphorylation and fetuin processing. One proposed functional role of this connecting peptide is in an insulin receptor inhibitory function (INSR) (59, 60). The INSR inhibitory activity of matured human fetuin isolated from plasma (assumed without the connecting peptide) is significantly lower compared with recombinant human fetuin (which is a single-chain protein including the connecting peptide). Because it has been reported that 7–20% of circulating fetuin carries phosphorylation at Ser330 (14), it is tempting to speculate that a similar portion of human fetuin circulates as a single-chain protein containing the connecting peptide (14, 57). The existence of such a minor, albeit metabolically more active pool of fetuin could be important in the pathology of type 2 diabetes which relates to acquired or inherited defects in the insulin receptor signaling cascade. Our data suggest a different scenario. We showed in our earlier study and here, that human serum fetuin exists exclusively in the two-chain form architecture, however it still contains the connecting

peptide. The proteolytic cleavage separating fetuin into two chains occurs at aforementioned position 340 and completely removes the Arg residue. Therefore, the modification of phosphorylation site Ser330 does not seem to correlate with the absence/presence of the connecting peptide and a putative regulatory function of the phosphate at Ser330 has likely a different mechanism.

**Sepsis Specific Alterations in the Proteoforms of Human Fetuin**—In many diseases, including acute and chronic inflammatory diseases, several acute phase proteins display altered glycosylation (61). Human fetuin has been classified as a negative acute phase glycoprotein, however, whether and how inflammatory processes affect fetuin glycosylation has not been described yet. Here, we observed in fetuin purified from septic patient donors a statistically significant increase of fucosylation on N-glycosylation site Asn176 (B). Notably, two septic patients, M7 and F7, did not show increased fucosylation. Both these patients were considerably younger compared with the other patients (18 and 19 years old). It has been previously hypothesized that N-glycan profiles are affected by age and gender (62–65). Although arguably on a small dataset, we did observe a correlation in our data between the extent of fucosylation and age. However, the number of data points showing the correlation between fucosylation and age was not enough to perform a proper statistical evaluation. Nonetheless, this further supports that the discovery and development of biomarkers should seriously take into consideration factors like gender and age of the donors. Whether the increased fucosylation has a functional consequence or is barely a diagnostic marker of “diseases-state” needs further exploration.

In conclusion, we analyzed here for the first time fetuin samples extracted from twenty individual donors, 10 healthy and 10 suffering from a septic shock. This personalized in-depth proteoform profiling provided a wealth of data on fetuin processing, fetuin phosphorylation and fetuin N- and O-glycosylation. One of the most intriguing findings of our study is that the common fetuin gene polymorphism affects the corresponding proteins’ proteoform profiles.

In most genomics and proteomics studies on fetuin so far consequences of the gene polymorphisms, and fetuin processing have been largely ignored. We feel that apparent discrepancies on fetuin reported in the literature, need to be readdressed, taking into account the donor’s individual genotype, physiological state, and fetuin processing state. It has been often suggested that fetuin is a protein sensitive to various physiological states, but whether its clinical potential is strong, deserves to be further investigated. Our data clearly demonstrates how a single gene can lead to a very broad range of proteoforms, which all may be functionally differentially active. At the proteoform level there seem to be no boring genes, as even fetuin, characterized already 75 years ago, still exposes new and exciting features when studied in

human individuals exhibiting different genotypes and physiological states.

**Acknowledgments**—We acknowledge support from the Netherlands Organization for Scientific Research (NWO) funding the large-scale proteomics facility *Proteins@Work* (project 184.032.201) embedded in the Netherlands Proteomics Centre. A.J.R.H. acknowledge further support by the NWO TOP-Punt Grant 718.015.003.

#### DATA AVAILABILITY

The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE (66) partner repository with the dataset identifier PXD012465 and 10.6019/PXD012465. As PRIDE does not fully support glycopeptide data Byonic results, in the form of freely available Byonic Preview files, containing all identified and assigned spectra are available at the following link: <https://figshare.com/s/c3ef7dba6c2079e0402f>.

☐ This article contains supplemental Figures and Tables.

|| To whom correspondence may be addressed. Tel.: +31302536149; E-mail: v.franc@uu.nl.

‡‡ To whom correspondence may be addressed. Tel.: +31302536797; E-mail: a.j.r.heck@uu.nl.

**Author contributions:** Y.-H.L., J.Z., and F.v. performed research; Y.-H.L., J.Z., A.B.M., and F.v. contributed new reagents/analytic tools; Y.-H.L., J.Z., F.v., and A.J.R.H. analyzed data; Y.-H.L., A.B.M., F.v., and A.J.R.H. wrote the paper; F.v. and A.J.R.H. designed research.

#### REFERENCES

1. Ix, J. H., Shlipak, M. G., Brandenburg, V. M., Ali, S., Ketteler, M., and Whooley, M. A. (2006) Association between human fetuin-A and the metabolic syndrome: Data from the heart and soul study. *Circulation* **113**, 1760–1767
2. Dziegielewska, K. M., Møllgård, K., Reynolds, M. L., and Saunders, N. R. (1987) A fetuin-related glycoprotein ( $\alpha$ 2HS) in human embryonic and fetal development. *Cell Tissue Res.* **248**, 33–41
3. Denecke, B., Gräber, S., Schäfer, C., Heiss, A., Wöltje, M., and Jahnen-Dechent, W. (2003) Tissue distribution and activity testing suggest a similar but not identical function of fetuin-B and fetuin-A. *Biochem. J.* **376**, 135–145
4. Cai, M. M. X., Smith, E. R., and Holt, S. G. (2015) The role of fetuin-A in mineral trafficking and deposition. *Bonekey Rep.* **4**, 1–10
5. Jahnen-dechent, LBW. (2013) The Role of Fetuin-A in Physiological and Pathological Mineralization. *Calcif. Tissue Int.* **93**, 355–364
6. Towler, D. A., Editor, G., Jahnen-dechent, Heiss, W. A., Scha, C., and Ketteler, M. (2011) Fetuin-A regulation of calcified matrix metabolism. *Circ. Res.* **108**, 1494–1509
7. Weikert, C., Stefan, N., Schulze, M. B., Pischon, T., Berger, K., Joost, H. G., Häring, H. U., Boeing, H., and Fritsche, A. (2008) Plasma fetuin-A levels and the risk of myocardial infarction and ischemic stroke. *Circulation* **118**, 2555–2562
8. Ketteler, M., Bongartz, P., Westenfeld, R., Wildberger, J. E., Mahnken, A. H., Böhm, R., Metzger, T., Wanner, C., Jahnen-Dechent, W., and Floege, J. (2003) Association of low fetuin-A (AHSR) concentrations in serum with cardiovascular mortality in patients on dialysis: A cross-sectional study. *Lancet* **361**, 827–833
9. Wang, H., and Sama, A. E. (2012) Anti-inflammatory role of fetuin-A in injury and infection. *Curr. Mol. Med.* **12**, 625–633
10. Yin, L., Cai, W. J., Zhu, L. Y., Li, J., Su, X. H., Wang, X. L., Chang, X. Y., and Sun, K. (2015) Association of plasma Fetuin-A and clinical characteristics in patients with new-onset type 2 diabetes mellitus. *Int. J. Clin. Exp. Med.* **8**, 991–999
11. Sun, Q., Cornelis, M. C., Manson, J. A. E., and Hu, F. B. (2013) Plasma levels of fetuin-A and hepatic enzymes and risk of type 2 diabetes in women in the U.S. *Diabetes* **62**, 49–55

12. Schäfer, C., Heiss, A., Schwarz, A., Westenfeld, R., Ketteler, M., Floege, J., Müller-esterl, Schinke, W. T., and Jahn-dechent, W. (2003) The serum protein alpha 2-Heremans-Schmidt glycoprotein fetuin-A is a systemically acting inhibitor of ectopic calcification. *J. Clin. Invest.* **112**, 357–366
13. Singh, M., Sharma, P. K., Garg, V. K., Mondal, S. C., Singh, A. K., and Kumar, N. (2012) Role of fetuin-A in atherosclerosis associated with diabetic patients. *J. Pharm. Pharmacol.* **64**, 1703–1706
14. Kaushik, S. V., Plaisance, E. P., Kim, T., Huang, E. Y., Mahurin, A. J., Grandjean, P. W. M. S. (2009) Extended-release niacin decreases serum fetuin-A concentrations in individuals with metabolic syndrome. *Diabetes. Metab. Res. Rev.* **25**, 427–434
15. Vörös, K., Gráf, L., Prohászka, Z., Gráf, L., Szenthe, P., Kaszás, E., Böröcz, Z., Cseh, K., and Kalabay, L. (2011) Serum fetuin-A in metabolic and inflammatory pathways in patients with myocardial infarction. *Eur. J. Clin. Invest.* **41**, 703–709
16. Ismail, N. A., Ragab, S., Abd El Dayem, S. M., Abd ElBaky, A., Salah, N., Hamed, M., Assal, H., and Koura, H. (2012) Fetuin-A levels in obesity: Differences in relation to metabolic syndrome and correlation with clinical and laboratory variables. *Arch. Med. Sci.* **8**, 826–833
17. Cox, D. W., Andrews, B. J., and Wills, D. E. (1986) Genetic polymorphism of alpha 2HS-glycoprotein. *Am. J. Hum. Genet.* **38**, 699–706
18. Osawa, M., Umetsu, K., Ohki, T., Nagasawa, T., Suzuki, T., and Takeichi, S. (1997) Molecular evidence for human alpha2-HS glycoprotein (AHSG) polymorphism. *Hum. Genet.* **99**, 18–21
19. Andersen, G., Burgdorf, K. S., Sparsø, T., and Borch-johnsen, K. (2008) AHSG tag single nucleotide polymorphisms associate with type 2 diabetes and dyslipidemia: studies of metabolic traits in 7,683 white Danish subjects. *Diabetes* **57**, 1427–1432
20. Kim, J. G., Kim, H., Ku, S. Y., Kim, S. H., Choi, Y. M., and Moon, S. Y. (2004) Association between human  $\alpha$  2-Heremans Schmidt glycoprotein (AHSG) polymorphism and endometriosis in Korean women. *Fertil. Steril.* **82**, 1497–1500
21. Brega, A., Scacchi, R., Cuccia, M., Kirdar, B., and Peloso, G. C. R. (1998) Study of 15 protein polymorphisms in a sample of the Turkish population. *Hum. Biol.* **70**, 715–728
22. Cerri, N., and De Ferrari, F. (1991) Genetic polymorphism of Alpha-2-HS glycoprotein in Lombardy (Italy). *Int. J. Legal Med.* **104**, 77–79
23. Elzanowski, A., Barker, W. C., Hunt, L. T., and Seibel-Ross, E. (1988) Cystatin domains in alpha-2-HS-glycoprotein and fetuin. *FEBS Lett.* **227**, 167–170
24. Araki, T., Yoshioka, Y., and Schmid, K. (1989) The position of the disulfide bonds in human plasma alpha 2 HS-glycoprotein and the repeating double disulfide bonds in the domain structure. *Biochim. Biophys. Acta* **994**, 195–199
25. Lin, Y. H., Franc, V., and Heck, A. J. R. (2018) Similar albeit not the same: in-depth analysis of proteoforms of human serum, bovine serum, and recombinant human fetuin. *J. Proteome Res.* **17**, 2861–2869
26. Hoffmann, M., Marx, K., Reichl, U., Wührer, M., and Rapp, E. (2016) Site-specific O-Glycosylation Analysis Of Human Blood Plasma Proteins. *Mol. Cell. Proteomics* **15**, 624–641
27. Spiro, R. G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R–56R
28. Seo, J., and Lee, K.-J. (2004) Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *J. Biochem. Mol. Biol.* **37**, 35–44
29. Smith, E. R., Ford, M. L., Tomlinson, L. A., Rocks, B. F., Rajkumar, C., and Holt, S. G. (2010) Poor agreement between commercial ELISAs for plasma fetuin-A: An effect of protein glycosylation? *Clin. Chim. Acta* **411**, 1367–1370
30. Kirwan, A., Utratna, M., O'Dwyer, M. E., Joshi, L., and Kilcoyne, M. (2015) Glycosylation-based serum biomarkers for cancer diagnostics and prognostics. *Biomed Res. Int.* **2015**, 16
31. Gornik, O., and Lauc, G. (2008) Glycosylation of serum proteins in inflammatory diseases. *Dis. Markers* **25**, 267–278
32. McCarthy, C., Saldova, R., Wormald, M. R., Rudd, P. M., McElvaney, N. G., and Reeves, E. P. (2014) 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
33. Flood-Nichols, S. K., Kazanjian, A. A., Tinnemore, D., Gafken, P. R., Ogata, Y., Napolitano, P. G., Stallings, J. D., and Ippolito, D. L. (2014) Aberrant glycosylation of plasma proteins in severe preeclampsia promotes monocyte adhesion. *Reprod. Sci.* **21**, 204–214
34. Brinkman-Van Der Linden, E. C. M., Van Ommen, E. C. R., and Van Dijk, W. (1996) Glycosylation of  $\alpha$ 1-acid glycoprotein in septic shock: Changes in degree of branching and in expression of sialyl LewisX groups. *Glycoconj. J.* **13**, 27–31
35. Brinkman-van der Linden, E. C., de Haan, P. F., Havenaar, E. C., van DW. (1998) Inflammation-induced expression of sialyl LewisX is not restricted to alpha1-acid glycoprotein but also occurs to a lesser extent on alpha1-antichymotrypsin and haptoglobin. *Glycoconj. J.* **15**, 177–182
36. Yang, Y., Liu, F., Franc, V., Halim, L. A., Schellekens, H., and Heck, A. J. R. (2016) Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* **7**, 13397
37. Franc, V., Zhu, J., and Heck, A. J. R. (2018) Comprehensive proteoform characterization of plasma complement component C8 $\alpha$  $\beta$  $\gamma$  by hybrid mass spectrometry approaches. *J. Am. Soc. Mass Spectrom.* **29**, 1099–1110
38. Franc, V., Yang, Y., and Heck, A. J. R. (2017) Proteoform profile mapping of the human serum complement component C9 revealing unexpected new features of N-, O-, and C-Glycosylation. *Anal. Chem.* **89**, 3483–3491
39. Havugimana, P. C., Wong, P., and Emili, A. (2007) Improved proteomic discovery by sample pre-fractionation using dual-column ion-exchange high performance liquid chromatography. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **847**, 54–61
40. Rose, R. J., Damoc, E., Denisov, E., Makarov, A., and Heck, A. J. R. (2012) High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods* **9**, 1084–1086
41. Rosati, S., Rose, R. J., Thompson, N. J., Van Duijn, E., Damoc, E., Denisov, E., Makarov, A., and Heck, A. J. R. (2012) Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angew. Chemie - Int. Ed.* **51**, 12992–12996
42. Bern, M., Caval, T., Kil, Y. J., Tang, W., Becker, C., Carlson, E., Kletter, D., Sen, K. I., Galy, N., Hagemans, D., Franc, V., and Heck, A. J. R. (2018) Parsimonious charge deconvolution for native mass spectrometry. *J. Proteome Res.* **17**, 1216–1226
43. Hart, G., Esko, J. D., Kinoshita, T., Cummings, R. D., Schnaar, R. L., Freeze, H. H., Etzler, M. E., Bertozzi, C. R., Lütteke, T., Dell, A., Vliegenthart, J. F., Stanley, P., Prestegard, J. J., Aebi, M., Rudd, P., Packer, N. H., Kanehisa, M., Kornfeld, S., Narimatsu, H., Bolton, E., Varki, A., Darvill, A., Paulson, J., Seeberger, P. H., Perez, S., York, W., Frank, M., Aoki-Kinoshita, K. F., Marth, J. D., Taniguchi, N., and Toukach, P. (2015) Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323–1324
44. Frese, C. K., Altalear, A. F. M., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* **10**, 2377–2388
45. Bern, M., Kil, Y. J., and Becker, C. (2012) Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinforma.*, Chapter 13: Unit13.20
46. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968
47. Čaval, T., Tian, W., Yang, Z., Clausen, H., and Heck, A. J. R. (2018) Direct quality control of glycoengineered erythropoietin variants. *Nat. Commun.* **9**, 3342
48. Karabudak, A., Giama, N. H., Moser, C. D., Miyoshi, E., and Roberts, L. R. (2017) Identification of fucosylated Fetuin-A as a potential biomarker for cholangiocarcinoma. *Proteomics - Clin. Appl.* **11**, 10
49. King, S. L., Joshi, H. J., Schjoldager, K. T., Halim, A., Madsen, T. D., Dziegiel, M. H., Woetmann, A., Vakhrushev, S. Y., and Wandall, H. H. (2017) Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells. *Blood Adv.* **1**, 429–442
50. Jentoft, N. (1990) Why are proteins O-glycosylated? *Trends Biochem. Sci.* **15**, 291–294
51. Ding, Q., Yang, L., Dinarvand, P., Wang, X., and Rezaie, A. R. (2015) Protein C Thr315Ala variant results in gain of function but manifests as type II deficiency in diagnostic assays. *Blood* **125**, 2428–2434
52. Grasbon-Frodl, E., Lorenz, H., Mann, U., Nitsch, R. M., Windl, O., and Kretzschmar, H. A. (2004) Loss of glycosylation associated with the

- T183A mutation in human prion disease. *Acta Neuropathol.* **108**, 476–484
53. Li, S., Iakoucheva, L. M., Mooney, S. D., and Radivojac, P. (2010) Loss of post-translational modification sites in disease. *Pacific Symp. Biocomput.* 337–347
54. Han, G., Ye, M., Zhou, H., Jiang, X., Feng, S., Jiang, X., Tian, R., Wan, D., Zou, H., and Gu, J. (2008) Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics* **8**, 1346–1361
55. Oppermann, F. S., Gnad, F., Olsen, J. V., Hornberger, R., Greff, Z., Kéri, G., Mann, M., and Daub, H. (2009) Large-scale proteomics analysis of the human kinome. *Mol. Cell. Proteomics* **8**, 1751–1764
56. Zhou, W., Ross, M. M., Tessitore, A., Ornstein, D., VanMeter, A., Liotta, L. A., and Petricoin, E. F. (2009) An initial characterization of the serum phosphoproteome. *J. Proteome Res.* **8**, 5523–5531
57. Goustin, A. S., and Abou-Samra, A. B. (2011) The “thrifty” gene encoding Ahsg/Fetuin-A meets the insulin receptor: Insights into the mechanism of insulin resistance. *Cell. Signal.* **23**, 980–990
58. Nawratil, P., Lenzen, S., Kellermann, J., Haupt, H., Schinke, T., Mu, W., and Jahnen-dechent, W. (1996) Limited proteolysis of human alpha 2-HS glycoprotein/fetuin. *Biochemistry* **271**, 31735–31741
59. Dziegielewska, K., Brown, W., Casey, S., Christie, D., Foreman, R., Hill, R., and Saunders, N. (1990) The complete cDNA and amino acid sequence of bovine fetuin. *J. Biol. Chem.* **265**, 4354–4357
60. Srinivas, P. R., Goustin, A. S., and Grunberger, G. (1995) Baculoviral expression of a natural inhibitor of the human insulin receptor tyrosine kinase. *Biochem. Biophys. Res. Commun.* **208**, 879–885
61. Mariño, K., Saldova, R., Adamczyk, B., and Rudd, P. M. (2012) Changes in serum N-glycosylation profiles: Functional significance and potential for diagnostics. *Carbohydrate Chem.* **37**, 57–93
62. Ding, N., Sun, H. N., Sun, W., Qu, Y., Liu, X., Yao, Y., Liang, X., Chen, C. C., and Li, Y. (2011) Human serum N-glycan profiles are age and sex dependent. *Age Ageing* **40**, 568–575
63. Vanhooren, V., Dewaele, S., Libert, C., Engelborghs, S., De Deyn, P. P., Toussaint, O., Debacq-Chainiaux, F., Poulain, M., Glupczynski, Y., Franceschi, C., Jaspers, K., van der Pluijm, I., Hoeijmakers, J., and Chen, C. C. (2010) Serum N-glycan profile shift during human ageing. *Exp. Gerontol.* **45**, 738–743
64. Vanhooren, V., Laroy, W., Libert, C., and Chen, C. (2008) N-Glycan profiling in the study of human aging. *Biogerontology* **9**, 351–356
65. Vanhooren, V., Liu, X. E., Franceschi, C., Gao, C. F., Libert, C., Contreras, R., and Chen, C. (2009) N-glycan profiles as tools in diagnosis of hepatocellular carcinoma and prediction of healthy human ageing. *Mech. Ageing Dev.* **130**, 92–97
66. Vizcaíno, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–D456