A Robust and Versatile Automated Glycoanalytical Technology for Serum Antibodies and Acute Phase Proteins: Ovarian Cancer Case Study

Article *in* Molecular & Cellular Proteomics · August 2019

DOI: 10.1074/mcp.RA119.001531

CITATIONS
8

READS
206

8 authors, including:

Roisin O'Flaherty
National University of Ireland, Maynooth
40 PUBLICATIONS **482** CITATIONS

SEE PROFILE

Ian Walsh
Bioprocessing Technology Institute
62 PUBLICATIONS **2,050** CITATIONS

SEE PROFILE

Henning Stöckmann
Eli Lilly
73 PUBLICATIONS **1,415** CITATIONS

SEE PROFILE

Mol Cell Proteomics Papers in Press. Published on August 30, 2019 as Manuscript RA119.001531

A robust and versatile automated glycoanalytical technology for serum antibodies and acute phase proteins: ovarian cancer case study

Authors:

Róisín O'Flaherty[§*], Mohankumar Muniyappa[§], Ian Walsh[φ], Henning Stöckmann[§†], Mark Hilliard[§], Richard Hutson[β], Radka Saldova[§μ], Pauline M Rudd[§]

§ NIBRT GlycoScience Group, National Institute for Bioprocessing Research and Training, Fosters Avenue, Mount Merrion, Blackrock, Dublin 4, Ireland, A94X099.
φ Bioprocessing Technology Institute, Agency for Science, Technology and Research (ASTAR), 20 Biopolis Way, #06-01 Centros, Singapore 138668, Singapore.
†current address: Abbvie Inc.1, Discovery Chemistry and Technologies, 1 North Waukegan Road, North Chicago, IL 60064, United States.
β Cancer Research UK Clinical Centre at Leeds, St James' University Hospital, Leeds LS9 7TF, UK.
μ UCD School of Medicine, College of Health and Agricultural Science, University College Dublin, Belfield, Dublin 4, Ireland.
*Correspondence may be addressed to: Róisín O'Flaherty; Email: roisin.oflaherty@nibrt.ie; Tel 0035312158154; NIBRT GlycoScience Group, National Institute for Bioprocessing Research and Training, Fosters Avenue, Mount Merrion, Blackrock, Dublin 4, Ireland.

Abstract:

The direct association of the genome, transcriptome, metabolome, lipidome and proteome with the serum glycome has revealed systems of interconnected cellular pathways. The exact roles of individual glycoproteomes in the context of disease have yet to be elucidated. In a move towards personalized medicine, it is now becoming critical to understand disease pathogenesis, and the traits, stages, phenotypes and molecular features that accompany it, as the disruption of a whole system. To this end, we have developed an innovative technology on an automated platform, "GlycoSeqCap", which combines *N*-glycosylation data from six glycoproteins using a single source of human serum. Specifically, we multiplexed and optimised a successive serial capture and glycoanalysis of six purified glycoproteins, immunoglobulin G (IgG), immunoglobulin M (IgM), immunoglobulin A (IgA), transferrin (Trf), haptoglobin (Hpt) and alpha-1-antitrypsin (A1AT), from 50µl of human serum. We provide the most comprehensive and in-depth glycan analysis of individual glycoproteins in a single source of human serum to date. To demonstrate the technological application in the context of a disease model, we performed a pilot study in an ovarian cancer cohort (n=34) using discrimination and classification analyses to

identify aberrant glycosylation. In our sample cohort, we exhibit improved selectivity and specificity over the currently used biomarker for ovarian cancer, CA-125, for early stage ovarian cancer. This technology will establish a new state-of-the-art strategy for the characterization of individual serum glycoproteomes as a diagnostic and monitoring tool which represents a major step towards understanding the changes that take place during disease.

Introduction:

Glycosylation is the most common and complex type of post-translational modification and glycoscience is rightly recognised as a current frontier. Over 1% of total human genome codes for approximately 300 functional glycogenes known to exist in humans[1]. Post-translational modifications (PTMs) of a protein are critical late events in protein biosynthesis and glycosylation is one of the most important. It reflects the genome of the individual person and also the many other influences on the cellular pathways. In fact, the glycoproteome can be thought of as the final readout of the genome that confers function on the gene products. A range of human genetic disorders, including congenital disorders of glycosylation[2,3], MODY type diabetes[4], galactosemia[5] and muscular dystrophies[6] have been directly linked to or shown to involve faulty glycosylation. Cancer associated aberrant gene regulation also results in alterations of glycan structures and has been well studied both in the serum glycome and specific serum glycoproteins[7-9]. In addition, chronic inflammatory diseases display altered glycosylation[10].

Immunoglobulin G (IgG) *N*-glycome is well characterised using liquid chromatography and mass spectrometry methods,[11-13] but *N*-glycosylation analysis of other glycoproteins has been studied to a far lesser extent. For example, *N*-glycosylation analysis of acute phase proteins, such as transferrin (Trf), alpha-1-antitrypsin (A1AT) and haptoglobin (Hpt), has been conducted in our laboratory[8,14] using complex methodologies such as isoelectric focussing (IEF) or 2D-gel electrophoresis for glycoprotein separation. In addition, *N*-glycosylation of serum immunoglobulin M (IgM)[15] and immunoglobulin A (IgA)[16] have been characterized previously using normal phase high performance liquid chromatography (NP-HPLC) a decade ago with only a proportion of the glycans identified, a limitation of the technology at the time. More recently, IgA site specific glycosylation (*N*- and *O*-glycopeptides from IgA1) was elucidated in serum of patients with rheumatoid arthritis (RA) during pregnancy using matrix-assisted laser/desorption ionisation Fourier transform ion cyclotron resonance

(MALDI-FTICR) mass spectrometry[17]. Of note, the team additionally performed released *N*-glycan analysis and could identify major *N*-glycans that were not identified in the characterized IgA1 *N*-glycopeptide fractions. As such, a detailed structural characterization of serum IgA *N*-glycome analysis remains to be undertaken. In a separate study, protein-specific differential glycosylation of immunoglobulins (IgG, IgM and IgA) were studied in serum of ovarian cancer patients[18] using multiple reaction monitoring on a triple quadrupole mass spectrometer. This technology presents similar limitations as to the total structural elucidation of the *N*-glycome, but is highly revealing, whereby the authors hypothesize that within the total serum glycome profile, which is largely dominated by the highest abundance proteins (such as antibodies IgG, IgM and IgA or acute phase proteins Trf, Hpt, A1AT) that protein- and site-specific glycosylation profiles will be likely to provide further insights into protein specific alterations in glycosylation of the glycans related to ovarian cancer as well as serve as more specific biomarkers for ovarian cancer than current tools[18].

Although some advances have been made in the sample purification and target glycoprotein enrichment[19], there is a growing necessity for the development of more robust, consistent, sensitive and versatile methods that can be automated and applied to personalized medicine approaches. To address this need, we optimised and automated a glycoanalytical technology to capture and glycoprofile six abundant individual glycoproteins by serial extraction-IgG, IgM, IgA, Trf, Hpt and A1AT. We performed a detailed analysis of the *N*-glycomes from these six glycoproteins using 50 µl of pooled normal human serum (NHS). Then we applied this technology to an ovarian cancer patient cohort, consisting of 7 healthy controls, 6 borderline and 21 metastatic ovarian cancer patients and tested its potential to be used as a diagnostic tool for earlier detection of ovarian cancer. This study is a follow-up of our previously reported sample preparation and chromatography technologies for glycan analysis[11,20,21] and biomarker discovery.

Experimental:

Serum samples

Normal Human Serum (NHS) samples were used as described previously[20]. Briefly human adult serum sample from healthy male and female blood donors were pooled and used as a source of glycoproteins for subsequent glycan analysis (courtesy of the U.K. Blood Transfusion Service). Serum samples from 7 healthy women (hereafter

called normal) and 27 ovarian cancer patients classified as either metastatic (n=21) or borderline (n=6) were used for the GlycoSeqCap technology (Supporting Tables S9) following ethical approval and obtaining informed consent. After allowing the blood to clot for 30-60min, serum was obtained by centrifugation at 2,000g for 10min and stored at -80^{o}C until analysis.

Chemicals and Reagents

All chemical reagents and solvents were purchased from Sigma-Aldrich. Pre-packed Protein G (PTH93-20-02) 300µL tip columns obtained from PhyNexus Inc. Albumin (Cat 191297005), IgM (Cat# 289005), IgA(Cat# 190288005), Trf (Cat# 191306005) and A1AT (Cat# 191287005) affinity matrix resins were obtained from Life Technologies CaptureSelect ^{TM} and Hpt (Cat# 291.2950.05) obtained from BAC BV, Capture Select. Exoglycosidase and PNGase F enzymes were obtained from NEB: PNGase F (P0709L), Bovine Kidney Fucosidase (BKF, P0748S), and other exoglycosidases *Arthrobacteria* sialidase (ABS, PZGK80090), Bovine Testes Galactosidase (BTG, PZGKX-5013), *N*-Acetyl Hexosaminidase (GUH, PZGK80050), and Jack Bean Mannosidase (JBM, GKX-5003) obtained from Europa (Prozyme). Solid-supported Ultralink hydrazide (Cat# 11809410) and Fermentas PageRuler Pre-stained Protein Ladder (Cat# 11832124) was obtained from Fisher Scientific. The following plates were used;1µm 96 well (PALL Acroprep Advance 350, VWR 518-0026), 96 well 2mL collection plate (X50 MICROPLATE, Fisher, 11511963), 96 well assay (Greiner, 450µl, Cruinn 731-1372), 384 well ultrafiltration (Acroprep, 10KDa, PALL 5077), 384 well PCR (Fisher, Armadillo AB-2384/O, orange, 30µl, VWR 732 1578) and a 384 well storage/collection (Corning, 240µl VWR 736 0202 3347). Reagents were placed on the robotic platform in a 12 well reagent trough (VWR 732-1390). Samples were prepared on a Hamilton robotics StarLet liquid-handling platform. The instrument is equipped with eight software-controlled pipettes, a vacuum manifold, and an automated heater shaker. Samples were analysed on a Waters Acquity H-class UPLC instrument. The following workflow was implemented as robotic program on Hamilton Robotics Venus One software.

Custom Phytip Glycoprotein affinity matrix resin tip columns preparation

Each specific anti-protein and anti-glycoprotein matrix was packed in Phytips by PhyNexus Inc. Briefly, 20µL each of Albumin, IgG, IgM, IgA, Trf, and A1AT affinity

matrix resins purchased from Life Technologies CaptureSelect ™ and Hpt purchased from BAC BV were packed individually in a 300µL Hamilton tips.

Phytip Glycoprotein Affinity Purification

Glycoproteins from 50µL of whole serum sample were captured in the following sequence: Albumin, Trf, IgG, IgM, IgA, Hpt and A1AT on the automated liquid handling station (Hamilton Starlet) in a 96-well format using the custom Phytips (Figure 1). In this method Phytip equilibration, capture, wash and elution refers to cycling a solution through the resin bed for a fixed number of aspirations and dispense steps termed cycles at defined flow rates. In a sequential fashion Albumin and the glycoproteins were captured from serum or from serum depleted of the preceding glycoproteins. For example, in the case of Trf, the Albumin depleted serum was used for the affinity chromatography. In the case of IgG affinity, the serum was depleted of both Albumin and Trf prior to IgG purification.

Albumin and the glycoproteins were captured using Phytips from serum/depleted serum in 96-well Greiner plates (50µL per well, 100µL mixing volume, 6 cycles, 20µL/s) using pre-equilibrated Phytips (200µL per well, 0.1M sodium phosphate buffer, 0.15M NaCl, pH 7.4, 3 cycles, 5µL/s). The Albumin/glycoprotein in Phytip resin were washed three times in 96 well Collection plates, each containing binding buffer (170µL per well mixing volume, 0.1M sodium phosphate buffer, 0.15M NaCl, pH 7.4, 6 cycles, 5µL/s) followed by treatment with washing buffer (170 µL per well mixing volume, 1% NaCl + 1g/L $NaN_{3,}$, 4 cycles, 5µL/s). The Albumin/glycoproteins were eluted (80µL per well, 0.2M Glycine-HCl buffer, pH 2.5, 3 cycles, 4µL/s) into three Greiner plates. Neutralisation buffer (10µL per well, 1M Tris-HCl buffer, pH 9.0) was added to two Greiner plates and the samples were pooled together for two Greiner plates (180µL resulting solution per well, mean concentrations 0.22, 0.90, 0.85, 0.12, 0.81 and 0.79mg/mL respectively for IgG, IgM, IgA, Trf, Hpt and A1AT).

1D SDS-PAGE

Affinity purified glycoproteins were reduced and loaded on SDS-PAGE gels (NuPAGE® Novex 4–12% Bis–Tris) and separated in a XCell SureLock™ Mini-Cell (Invitrogen, Carlsbad, CA) for 90min at 120V using a MES running buffer. 5µg of total protein was loaded in each lane of the gel. Once electrophoretic separation was

completed, proteins were visualized by staining with Coomassie blue staining solution followed by destaining with multiple changes of deionised water (Figure 1).

Automated glycoprotein denaturation and *N*-Glycan release

As described previously[22], briefly *N*-glycan analysis was performed using a pooled serum sample from 100 apparently healthy male and female adult blood donors (U.K. Blood Transfusion Service). The antibodies IgG, IgA, IgM and the acute phase proteins Trf, Hpt and A1AT were purified from the pool using the respective affinity resin (Life Technologies) packed in Phynexus Phytips as previously described (in above section). On an automated platform, the glycoprotein samples were dispensed into two 384-well ultrafiltration plates (maximum loading: 60µg protein, 10kDa). Ultra-filtration was performed by centrifugation (3700g, 30min, room temperature) or on a vacuum manifold (> 25 in Hg vacuum, 30min). Denaturation buffer (25µL per well, 100mM sodium bicarbonate, 50mM dithiothreitol (DTT), 0.1% sodium dodecyl sulfate (SDS)) was dispensed into 2 × 384 well ultrafiltration plates. After 10min incubation at room temperature, the samples were mixed 10 times (mixing volume: 15µL, flow rate: 10µL/s) and transferred to a 384-well PCR plate (Armadillo). The plate was placed into a robotic incubation chamber at 95°C for 10min. The plate was removed from the incubator and equilibrated to room temperature for 10min. 1M iodoacetamide (IAA), 10µL, was dispensed into each well of the ultrafiltration plate and the samples (25µl) were transferred back into the 384 well PCR plate. The samples were mixed 5 times (mixing volume 20µL, flow rate, 10µL/s). After 10min incubation at room temperature, the ultrafiltration plate was stacked onto a 240µL collection plate (Corning block) and centrifuged (3700g, 30min, room temperature) and the supernatant was removed. Next, 10µL of a 25mM sodium bi-carbonate solution was dispensed into each well and the ultra-filtration plate was centrifuged (3700g, 30min, room temperature) and the supernatant was removed.

The deglycosylation mix, 12µL (0.4µl of PNGase F (2.5U/mL), 25mM sodium bicarbonate) was dispensed into each well and the plate was then covered with a lid and incubated on a robotic orbital shaker (shaking orbit: 2mm, shaking speed: 700rpm, temperature: 38°C, incubation time: 30min). The ultrafiltration plate was stacked onto a 30 µL/well PCR plate (Armadillo) and centrifuged (3700g, 10min, room temperature). Finally, 10µL of 25M sodium bicarbonate solution was dispensed into each well of the ultrafiltration plate (still stacked onto the PCR plate) and the assembly was centrifuged

(3700g, 10min, room temperature). The released *N*-glycans in the PCR plate were subsequently fluorescently labelled. For glycan labelling, 5µL of glycan sample (PCR plate) was transferred to a 95µL Corning block and 11.6µL of AQC (3mg/mL MeCN) was added. 3µL of this crude mixture was directly injected into the UPLC system. Alternatively, after the PNGase F release, glycans can be frozen and are stable at -20°C for labelling later.

Ultra-Performance Liquid Chromatography (UPLC)

As previously described[22], briefly the separation of AQC-derivatized *N*-glycans was carried out by UPLC with fluorescence detection on a Waters ACQUITY UPLC H-Class instrument consisting of a binary solvent manager, sample manager, and fluorescence detector under the control of Empower 3 software (Waters, Milford, MA, USA). The HILIC separations were performed using a Waters Ethylene Bridged Hybrid (BEH) Glycan column (150 × 2.1mm i.d., 1.7 µm particles) with 50mM ammonium formate (pH 4.4) as solvent A and MeCN as solvent B. The column was fitted with an ACQUITY in-line 0.2µm filter. The separation was performed using a linear gradient of 70−53% MCN 0.56mL/min in 16.5min for IgG *N*-glycan separation. An injection volume of 3µL prepared in 70% v/v MeCN was used throughout. Samples were maintained at 5°C prior to injection, and the separation temperature was 40°C. The FLD excitation /emission wavelength were λex = 245nm and λem = 395nm, respectively. The system was calibrated using an external standard of hydrolyzed and 2-AB-labeled glucose oligomers to create a dextran ladder, as described previously[11]. A fifth-order polynomial distribution curve was fitted to the dextran ladder to assign glucose unit (GU) values from retention times (using Empower software from Waters).

Liquid Chromatography-Mass Spectrometry (LC-MS)

Online coupled fluorescence (FLR)-mass spectrometry detection was performed using a Waters Xevo G2 QTof with Acquity® UPLC (Waters Corporation, Milford, MA, USA) and BEH Glycan column (2.1 x 150mm, 1.7µm particle size). For MS acquisition data the instrument was operated in positive-sensitivity mode with a capillary voltage of 3kV. The ion source block and nitrogen desolvation gas temperatures were set at 120°C and 350°C, respectively. The desolvation gas was set to a flow rate of 800L/h. The cone voltage was maintained at 40V. Full-scan data for glycans were acquired over m/z range of 300 to 2000. Data collection and processing were controlled by

MassLynx 4.1 software (Waters Corporation, Milford, MA, USA). The fluorescence detector settings were as follows: λexcitation: 245nm, λemission: 395nm; data rate was 10pts/second and a PMT gain = 20. Sample injection volume was 10µL (75% MeCN). The flow rate was 0.400mL/min (unless specified) and column temperature was maintained at 60°C; solvent A was 50mM ammonium formate (pH 4.4) and solvent B was MeCN. A 60min linear gradient was used and was as follows: 25-46% A for 35min, 46-80% A for 8min (flow rate at 0.2mL/min), 80-25% A for 27min. To avoid contamination of system, flow was sent to waste for the first 1.2min and after 55min.

Exoglycosidase digestions

The AQC labelled *N*-glycans from the glycoproteins (IgG, IgM, IgA, Trf, Hpt and A1AT) purified from healthy human serum were treated with exoglycosidases according to the literature procedure[23]. For enzymes ABS (Arthrobacter ureafaciens sialidase), BTG (Bovine testes beta-galactosidase), GUH (β-*N*-(1-2,3,4,6) Acetylglucosaminidase S), BKF (Bovine kidney alpha-fucosidase) and JBM (α1-2,3,6 Mannosidase J, Jack Bean): 1, 2, 2, 4 and 4µl of each was used per digestion to give the final concentrations of 0.5, 1, 8, 3200 and 60U/mL respectively. All digestions were carried out in a final volume of 10µl, at 50mM NaOAc pH 5.5 for 24h-96h.

Computational procedures

Statistical analysis: Variables age, menopause status, logCA125, C-reactive protein levels (CRP), protein titre and glycan peak areas between patients and controls were compared using a Tukey honest significant difference (HSD) test with ANOVA. Peak areas are compositional data (presented as a % of the total area under the graph) and therefore the constant-sum constraint (CSC) occurs. The CSC means individual variables do not vary independently, violating common assumptions upon which standard statistical analyses are performed. This was avoided by performing a log transform, log(Peaki)/(100-Peaki), on all peak areas[24]. To calculate the p-values the data was controlled for age confounder by creating a linear regression model (ANCOVA). Correction for multiple testing was performed using Benjamini-Hochberg procedure with a 10% false positive rate.

Missing values: There was one individual with missing age, menopause status, CA125, CRP and protein status. One individual had a missing menopause status. The small number of missing values were imputed using the K nearest neighbour approach

where missing attributes were predicted. This is an acceptable procedure when the number of missing values are small.[25]

Classification model: To classify individuals into normal, borderline or metastatic neural networks without hidden layers were trained with combinations of age, menopause status, logCA125, CRP, protein and glycan peak relative abundances. A one against all binary classification was performed (e.g. normal vs. borderline+metastatic). In order to test the performance of the models the data used to optimize parameters (train set) was split and a 10-fold cross validation (90% of data for training and 10% of data to test the model repeated 10 times) was performed. Discriminative power was calculated by Area under the Receiver Operating Characteristic (ROC) curve (AUC), sensitivity and specificity.

Clustering: Principal component analysis was performed using variables peaks areas, age, logCA125, CRP, protein titre and menopause status. For hierarchical clustering of the glycomic profiles, GP % areas were normalized to the range [0,1] and a hierarchy of clusters were built using the Ward algorithm[26]. Spearman correlations were used to calculate the similarity among GP, age, logCA125, CRP and protein titre. Therefore, correlated variables appear close together in the hieratical dendrogram produced by the clustering.

Results and Discussion:

GlycoSeqCap- automated *N*-glycan analysis platform using serial capture of individual serum glycoproteins

We developed a technology termed “GlycoSeqCap” involving the serial capture and subsequent *N*-glycoprofiling of individual glycoproteins from human serum. The multiplexed automated serial capture of six abundant serum glycoproteins (the immunoglobulins, IgG, IgM, IgA, and the acute phase proteins, transferrin (Trf), alpha-1-anti-trypsin (A1AT), haptoglobin (Hpt)) and *N*-glycan release from pooled human serum was optimised using a 96 well format on a robotic platform **(**Figure 1). Specific anti-glycoprotein capture resins (containing antibodies against specific glycoproteins) were packed in a series of PhyNexus phytips and each serum sample was passed successively through each resin in a sequential order Trf, IgG, IgM, IgA, Hpt and A1AT. The purified proteins, >98% pure as determined by 1D-SDS page, were then subjected

to PNGaseF release, ultrafiltration and aminoquinoline carbamate (AQC) fluorescent labelling of the *N*-glycan pools using the Hamilton Starlet robotic system using a previously published protocol (Figure 1)[11]. These pools were subjected to hydrophilic ultra-high performance liquid chromatography (HILIC-UPLC) followed by exoglycosidase array glycan analysis according to the established procedures in our laboratory[23]. The UPLC chromatograms for the fluorescently labelled released *N*-glycans of affinity purified glycoproteins (IgG, IgM, IgA, Trf, Hpt and A1AT) from human serum and the corresponding released *N*-glycans from total human serum[21] are presented with annotations for the major *N*-glycans (Figure 2).

Advantages and limitations of GlycoSeqCap technology

This technological advancement provides *N*-glycosylation information for IgG, IgM, IgA, Trf, Hpt and A1AT and serves as a template for many other glycoproteins of interest which may find use as a personalized medicine tool for future applications. A major advantage of the developed workflow is the ability to have multiplexed glycomics information from a single clinical source. In addition, the Phytips are reusable and show good reproducibility over three consecutive runs and after 1 month of storage. However, one limitation regarding the extension of the technology to include other glycoproteins is the requirement for an affinity re

sin with high selectivity and specificity for the targeted glycoprotein (e.g. antibodies or lectins) and in certain cases high amounts of biomaterial (in this case serum) may be required for lower abundant glycoproteins. In addition, the analyst must be cognisant of the purity of the targeted glycoprotein prior to enzymatic release of the *N*-glycans. For example, a recent publication cautions the users to account for the ubiquitous presence of varying levels of other contaminating plasma glycoproteins [27]. In our study, we were careful to assess the purity of the glycoproteins by SDS-PAGE (Figure 1) and systematically altered the number of capture, binding and wash steps to maximise purity prior to PNGase F release of *N*-glycans. As in the case of any affinity purification, it is possible that small traces of other glycoproteins have contributed to the *N*-glycan structures identified, albeit in tiny proportions.

The glycosylation data for six abundant glycoproteins in normal human serum aligns nicely with the total glycan pool identified for human serum previously (Figure 2)[21,28]. All major glycans from human serum are identified in one/more of the individual

glycoproteins that were characterized. In a large part, this identification on an individual glycoprotein level was enabled on such a small amount of serum (50µL) due to the use of a fluorescent label AQC which shows a twenty fold increase in fluorescent detection compared to the traditional label, 2-AB for released *N*-glycans which we have previously described for human IgG previously[11,23]. Various high-throughput large-scale studies have been conducted on a serum *N*-glycome level since 2009[29-31]. These studies have yielded significant contributions to understanding the role of glycosylation in many diseases but cannot pinpoint the exact glycosylation processing pathways involved. On an individual glycoprotein level, large scale studies have focussed on IgG *N*-glycome only to date,[32] leaving plenty of scope for future investigations on an individual glycoprotein level. In addition, it may also be possible to link significant alterations from the serum *N*-glycome to a specific glycoprotein by extrapolation using the major glycans identified in this work.

Profiling and detailed *N*-glycan analysis of serum IgG, IgA and IgM

In order to execute glycoprofiling to compare glycosylation of two distinct populations (such as serum IgG *N*-glycosylation from a normal population compared to an ovarian cancer cohort) it is important to generate reproducible profiles and fully characterize the *N*-glycomes. To this end, we have undertaken the most comprehensive study of released *N*-glycans for individual glycoproteins in human serum to date. UPLC chromatograms (for each individual glycoprotein) were integrated and split into glycan peaks (GPs) and each GP typically accounts for one or more glycans.

The serum IgG *N*-glycome UPLC chromatogram was split into 25 GPs (adapted from a literature procedure using 23 GPs[11]) to account for *N*-glycans prominent in IgG human serum from patients with ovarian cancer. The 25 IgG GP areas (G1-G25) were plotted for five technical replicates over three different days to provide standard error bars on the integrated peak areas (Figure 3a, Supporting Table S7). The majority of the GPs (21/25) are strongly reproducible for the technical replicates with coefficient of variance (CV) values for each GP below 25% with the exceptions of GPs G1, G15 and G25. These GPs cumulatively account for a small amount (0.35%) of the total peak area of IgG *N*-glycome in the technical replicates. Characterization proceeded as follows: the preliminary structures, assigned from glucose unit (GU) values (obtained by matching elution positions to a standard dextran hydrolysate curve and

using Glycobase software (data now migrated to Glycostore (https://www.glycostore.org/)[33]), were confirmed by exoglycosidase array digestions (Figure 3b) and intact mass using electrospray (ESI-LC) liquid chromatography mass spectrometry (Supporting Table S1). They are in agreement with literature[11,32] (with the addition of a high mannose species (M7) not previously identified) and are presented in Supporting Table S6.

The equivalent data for serum IgM and IgA *N*-glycomes were split into 24 (M1-M24) and 25 (A1-A25) GPs respectively and are shown in Figure 3 (reproducibility data in Supporting Table S7). For IgM, all GPs (M1-M24) show good reproducibility for the technical replicates (below 25%). Similarly IgA shows good reproducibility for all major GPs (A2-A25, below 25%) but exhibits slightly higher CVs for the first and last GPs (A1, A25) accounting for 0.92% of the total IgG *N*-glycome. Structural assignments of IgM and IgA *N*-glycans were characterised in a similar manner to serum IgG using GU values and LC-MS (Supporting Table S2 and S3) and are presented in Supporting Table S6. For serum IgM, structural assignment of *N*-glycans were in agreement with the previously reported literature values[15] with some additional glycans identified in our analysis (13 newly characterized compositions out of a total of 53) including some of the M5A1 Series (M5A1, M5A1G1 and M5A1G1S(6)1), FA1 Series (FA1, FA1G1 and FA1G1S1), A2 Series (A2G1S(6)1 and A2G2S(6,6)2) and the A2B Series (A2BG1S(6)1). For IgA, structural assignment of *N*-glycans were in agreement with the literature reports previously described[16,17,34]. As in the case of IgM, additional glycans were identified (16/55) in our in-depth analysis not previously described in the literature including some monoantennary and biantennary species (A1, A2[6]G1, A2[3]G1, FA2), hybrid species (M4A1, M4A1G1, M4A1G1S(6)1, M4A1BG1, M5A1G1, M5A1G1S(6)1), A3 Series (A3, A3G2, A3G2S(6)1, A3G3S2, FA3) and M10.

Profiling and detailed *N*-glycan analysis of serum Trf, Hpt and A1AT

The acute phase proteins, Trf, Hpt and A1AT were purified from human serum by serial extraction as described earlier and the released *N*-glycans labelled with AQC analysed by UPLC-HILIC. The 28 Trf (T1-T28), 31 Hpt (H1-H31) and 28 A1AT (AT1-AT28) glycan peak areas were plotted for at least five technical replicates over three different days to provide the standard error bars (Figure 4a,b and reproducibility data in Supporting Table S7). All GPs in Trf, Hpt show good reproducibility (CV below 25%)

except GPs T1 and T5 (total glycan peak area of 0.03%) in Trf and H5 (glycan peak area of 0.02%) in Hpt, accounting for very minor constituents of the total *N*-glycome in the technical replicates. The reproducibility of A1AT *N*-glycome is the lowest (being the last glycoprotein to be purified in the sequence) with only 19/28 glycan peaks exhibiting good reproducible CV values below 25%. The remainder (AT1, AT3-5, AT8-10, AT12, AT14) accounts for a total of 2.67% of the total A1AT *N*-glycome, a relatively small proportion of the total *N*-glycome.

The AQC labelled Trf, Hpt and A1AT (Figure 4) *N*-glycan pools were sequenced using exoglycosidase arrays and the data combined with glucose unit (GU) values and LC-MS to facilitate glycan identification. As for the immunoglobulins, many newly identified *N*-glycans are presented as well as agreement with the major glycans already presented in the literature for Trf [8],Hpt[8,35] and A1AT[8,36]. The list of newly identified glycans are highlighted with a Δ symbol in Supporting Tables S6. For Trf, 28 out of 48 glycan compositions are identified for the first time in this publication. Most significantly, we identify the FA3G3 series not identified for Trf previously which accounts for almost 4% of the total *N*-glycome. Similarly, for Hpt and A1AT, 32/47 and 28/44 glycan compositions are identified respectively for the first time. For Hpt, the FA3 series is significant accounting for approx. 3% of the total *N*-glycome and the newly identified glycan A3F1G3S3 accounts for approx. 2% of the peak area.

Comparison of glycosylation between antibody classes: IgG, IgM and IgA

We probed the glycosylation of selected antibodies to gain an insight into the structure-function relationship of antibody classes. One important point to consider is that serum glycoproteins are a combination of active components and waste products and this may complicate interpretations. Structural similarities and differences in antibody glycosylation of IgG, IgM and IgA are presented for pooled normal human serum (Figure 5a,b). The derived glycosylation traits were generated from UPLC data (Supporting Tables S6) and calculated from the summation of individual GPs for each glycoprotein (Supporting Table S8). Striking variation in glycosylation traits such as galactosylation, fucosylation, sialylation, bisecting GlcNAcs (glycans with bisecting *N*-acetylglucosamine residues), high mannose, hybrid or the degree of branching (monoantennary, biantennary, triantennary) can be observed. Overall, all antibodies exhibit a high proportion of biantennary structures (>60%) with little/no

monoantennary/triantennary structures, as documented previously in the literature[15,32,34]. Additionally, a high degree of galactosylation is observed for all three glycoproteins (>60%). Only α2,6-linked sialic acids (containing the Neu5Ac form) and no corresponding α2,3-linked sialic acids are identified for IgG or IgM which is consistent with literature reports[15,32]. Additionally, only minor species contained α2,3-linked sialic acids in IgA, as identified previously[17]. Increased sialylation was observed from IgG to IgM to IgA (22%, 60% and 85% respectively). IgG contains the highest proportion of core fucosylation (93%) relative to IgM or IgA (58% and 28% respectively) and no outer arm fucose were identified for the serum antibody series. This observation is consistent with other reports for serum antibodies[15,17,32]. IgA contains the highest abundances of bisecting GlcNAcs (52%) and galactose residues (93%) and a very small proportion of triantennary structures (0.9%) not observed for the other antibodies. IgM exhibits the lowest amount of biantennary structures (64%) and galactosylation (65%) but significantly higher amounts of high mannose structures (31%).

Acute phase protein glycosylation traits: Trf, Hpt and A1AT

Unlike the selected antibodies which share structural similarities in terms of their protein structure, the selected acute phase proteins (Trf, Hpt and A1AT) have less commonality and exhibit more varied glycosylation traits. These structural similarities and differences are presented for pooled normal human serum (Figure 5a,c). As for the antibodies, a high degree of galactosylation was observed (>95%) for selected acute phase proteins but relatively low amounts of total fucosylation (<20%) were identified, accounting for a combination of core fucose and the outer arm fucose residues (not identified in the selected serum antibodies in this study). Sialyl Lewis^x (SLex) motifs (Neu5Acα2-3Galβ1-4[Fucα1-3]GlcNAcβ), which contain outer arm fucose residues, are identified in human serum acute phase proteins, albeit in small proportions. Notably, as in the case for the antibody glycosylation, α2,6-linked sialic acids (>65%) are dominant over the corresponding α2,3-linked sialic acids (<10%) of the annotated sialic acids. In previous studies of acute phase proteins, many sialic acid linkage types were left ambiguous[8,35], as such we provide the most extensive breakdown to date for acute phase protein *N*-glycosylation. A larger proportion of higher branched structures are observed such as triantennary structures (>10%) and the presence of tetraantennary glycans are observed. Hpt exhibits the highest relative

amount of both triantennary (22%) and tetraantennary (6%) structures compared to Trf or A1AT.

Glycosylation traits for functional studies

The quantification of glycosylation traits for a series of glycoproteins is presented here for the first time from a single source of human serum (Figure 5). One can envisage this information as being a starting point for functional studies to investigate the specific role of glycosylation in our immune system and has the added advantage over traditional approaches in glycoanalytics to explore the effect of more than one glycoprotein in a single experiment. For example, little is known about how the most abundant antibody isotypes in human serum (IgG, IgM and IgA) compete for antigens but using quantified glycan traits, combined with antibody-antigen assays could provide a clearer picture for the role of glycosylation in this context. From our glycosylation traits alone, it is impossible to predict biological function but it may provide a basis for further investigations. For example, taking into consideration functional examples of afucosylation of IgG1 increasing antibody-dependent cellular cytotoxicity (ADCC)[37,38], and looking at levels of antibody fucosylation in human serum, the striking difference between IgG (93%) and IgM or IgA core fucosylation (58% and 28% respectively) may hint that serum IgA recruit effector cells for ADCC more efficiently compared to IgG or IgM. Importantly, it must be considered that we have measured only in human serum samples and this data does not necessarily reflect on a tissue/cell specific antibody glycosylation level. In addition to core fucosylation, we reported no outer arm fucose for the antibody classes. Noteworthy, outer arm fucosylation has been reported for secretory component (SC) *N*-glycans of IgA, but none for the corresponding J or H chain *N*-glycans[34]. It is known that serum IgA is predominantly monomeric and does not contain the J chain and the SC and as such we report no conflicting findings in this study. Another interesting differentiation between the glycosylation traits of the antibody isotypes is the high abundance of sialylation for IgA (85%) relative to IgM (60%) and IgG (22%). We propose that serum IgA uses electrostatic interactions and steric obstruction/adhesion as the main mechanistic mode for fighting infection but it is not clear if this sialylation is employed for antigen binding/effector function purposes. A more in-depth analysis for Fc and Fab *N*-glycosylation of serum IgA would shed more light in this regard. Lastly for the antibody class, we reported the highest proportion of high mannose structures (31%)

for serum IgM. This proportion of high mannose was higher than the value of 23% previously reported for serum IgM and the corresponding amount of galactosylation was lower (65%) compared to the literature value of 84%[15]. We suggest that advances in modern technology have allowed for a more accurate and precise assignment of *N*-glycans in our study- however it cannot be ignored that the human serum populations used in the two studies may have been different.

In contrast with immunoglobulins, which are mainly produced by B-cells, major plasma glycoproteins including Trf, Hpt and A1AT originate from hepatocytes, which express only very low levels of the FUT8 fucosyltransferase and thus contain a low percentage of core fucosylated glycans [32]. As anticipated, we observed lower levels of core fucose for the acute phase proteins in comparison to the antibody classes in our study (Figure 5). Another interesting point relates to the increased complexity of the acute phase protein *N*-glycosylation compared to immunoglobulin classes (Figure 5), with tetraantennary glycans present in these glycoproteins not isolated from the antibody classes. We cannot find a biological rationale for the reasons as to why the *N*-glycans are more branched in these acute phase proteins but it may be related to their inflammatory properties. Also, these glycoproteins contain SLex motifs. SLex are known to play a vital role in cell-cell recognition and other processes and are known to be overexpressed in cancer cells[39]. As such, this feature is particularly important in the context of our ovarian cancer cohort.

Glycoprofiling and statistical significance of selected glycoproteins in ovarian cancer

Having comprehensively assigned *N*-glycans and calculated derived glycosylation traits for the selected glycoproteins IgG, IgM, IgA, Trf, Hpt and A1AT in normal human serum, we glycoprofiled a cohort of patients with ovarian cancer to exploit the quantifiable *N*-glycosylation alterations for a) detection of ovarian cancer and b) to differentiate between stage of the disease in an effort to provide a clinical tool for early diagnosis in the future.

The *N*-glycome of selected proteins (IgG, IgM, IgA, Trf, Hpt and A1AT) from 7 normal controls and 27 ovarian cancer patients (either classified as borderline or metastatic presented in biological manifest in Supporting Table S9) as a test group were analysed

resulting in a total of 204 processed glycoprofiles. The profiles were split into glycan peaks (GPs) according to the individual glycoproteins (e.g. G1-G25 for IgG) and the values are presented as a proportion of total % peak area in Supporting Tables S10-S15 for glycoproteins IgG, IgM, IgA, Trf, Hpt and A1AT respectively. The derived glycosylation traits were also calculated and were derived as described earlier (Supporting Table S8). Controlling for age, the statistical significance (Tukey honest significant difference (HSD) with ANOVA, normally distributed (data not shown)) was measured for all GPs and glycosylation traits for each selected glycoprotein and are presented in Supporting Tables S16-S21, as well as for the clinical variables (Supporting Table S22) between normal vs. borderline, metastatic vs. borderline, normal vs metastatic clinical samples. The p-values were corrected for multiple testing error, using a 5% false discovery rate (FDR) approach proposed by Benjamini-Hochberg[40]. An adjusted $p<0.05$ was considered statistically significant. Significant GPs, glycosylation traits and clinical parameters for the patients are presented in Figure 6 for the glycoprotein series. Boxplots (6c) and the major glycan (6b) for the statistically significant GPs/glycosylation traits are also presented.

From the glycoproteins selected, Trf shows the best discrimination between normal patients (n=7) and borderline (n=6) or metastatic patients (n=21). Most notably, Trf glycosylation can distinguish between normal and borderline samples using four individual GPs (T11, T16, T17 and T18), whereby CA125, the gold standard for ovarian cancer detection does not show the same statistical significant separation. Fucosylation also stands out as a feature that may be exploited for this discrimination. The GPs/glycosylation features of Trf cannot differentiate between metastatic and borderline samples in this cohort however, whereas CA125 does show a significance.

Several studies have presented evidence implicating Trf in ovarian cancer biology [41]. In addition, alterations in Trf glycosylation in the context of cancer and other inflammatory diseases have been reviewed in the literature [36,42]. Taken together, Trf sialylation is often altered in disease states and is no exception in this study, whereby sialic acid (α2,6) is altered in the metastatic cancer cohort compared to normal controls. Since desialylated Trf has faster clearance, it may be an evolutionary tactic of bacteria/pathogens to contribute to oncogenesis. What is more striking in our study is the alterations of Trf fucosylation. Trf fucosylation has been shown to be dysregulated in diseases such as classical galactosemia[43] but to the best of our

knowledge has not been reported in the context of ovarian cancer to date. Future studies are warranted to investigate further.

In addition, haptoglobin glycosylation is significantly altered in metastatic cancer (n=18) compared to normal patients (n=7) in this cohort by six distinct GPs (H2, H3, H11, H20, H21 and H22) as well as glycosylation features fucosylation and SLex motif. This data is consistent with literature reports[44], whereby the main glycosylation alterations of Hpt in cancer appear to be the presence of aberrantly fucosylated and sialylated structures as well as increased branching.

Discrimination and cluster analysis in ovarian cancer

A discrimination analysis was explored for the probabilistic classification of healthy (normal) vs ovarian cancer samples using the glycosylation data and comparing to CA125 values for the clinical cohort. CA125 antigen is a high molecular weight glycoprotein, which is expressed by a large proportion of epithelial ovarian cancers and is currently regarded as the golden standard for ovarian cancer diagnosis, despite its poor sensitivity and specificity. It is only raised in approximately 50% of stage 1 epithelial ovarian cancers and in 75–90% of patients with advanced disease and false positive results have been noted in many medical disorders, both malignant and benign[45]. The area under the ROC curve (AUC), sensitivity (SEN) and specificity (SPE) were calculated and associated with normal vs. borderline, metastatic vs borderline and normal vs. metastatic clinical samples (Supporting Table S23, S24 and S25 respectively). The individual models performed well with regards to distinguishing normal from borderline patients, with the most promising result using derived traits from Hpt (AUC=1.000, SEN=1.000 and SPE=1.000) (Figure 7a). This finding, if validated could have major ramifications for early detection of ovarian cancer using Hpt glycosylation. Similarly, good discrimination was observed for normal vs metastatic patients. Again Hpt showed perfect discrimination for all peaks, derived traits and combinations thereof. However, CA125 also allows perfect discrimination so this negates the glycome data in this instance. No glycosylation feature could discriminate between borderline and metastatic patients and superior results were observed for CA125 in this case.

Cluster analysis was performed using principal component analysis (PCA) and hierarchical clustering. PCA results are provided in Supporting Figures S7-S12, and

hierarchical clustering is presented in Supporting Figures S13-S18 for the respective glycoproteins IgG, IgM, IgA, Trf, Hpt and A1AT for individual glycan peaks. No clear discrimination was observed between normal, borderline or metastatic samples for PCA analysis on individual glycan peaks. For hierarchical clustering, no major clustering was observed for glycoproteins IgG, IgM or A1AT but clustering was observed for IgA, Trf and Hpt with clear clustering of metastatic and normal samples respectively with the most pronounced affect for Hpt. Taking glycosylation traits into consideration, Hpt again outperforms the other glycoproteins and PCA plots are shown in Figure 7b showing a clear separation for normal and borderline samples and normal and metastatic samples but no separation between borderline vs metastatic samples, consistent with the AUC, SEN and SPE observed for this dataset.

Glycoanalytical diagnostic tools for ovarian cancer

The availability of diagnostic tools for ovarian cancer remains somewhat elusive. CA125 is currently the best diagnostic tool for ovarian cancer on the market, but is not reliable for diagnosing early stage ovarian cancer[46,47]. Early stage detection and subsequent treatment of ovarian cancer is an attractive approach to reduce morbidity from ovarian cancer. In this study, we corroborate these literature findings and show that CA125 cannot differentiate between borderline (n=6) and normal (n=7) samples (Figure 6). For IgG, IgA, IgM, Hpt and A1AT, again we cannot see any significant alterations between these classes but find instead that Trf glycosylation can discriminate using GPs T11, T16, T17 and T18 (Figure 6) in our small clinical cohort. If these findings could be replicated and validated, Trf glycosylation could be exploited as an early detection tool for ovarian cancer. To probe the potential role of Trf glycosylation in ovarian cancer on a very basic level, we measured the upregulation/downregulation of the relative abundance of *N*-glycans for the statistically significant GPs (Figure 6b) and hypothesize that the dominant glycosylation features may be significant in the progression of the disease. Two of the major glycans that are upregulated in the borderline samples (n=6) contain core fucose- FA2G2 (T11) and FA2G2S(6)1 (T16) whereas there is a corresponding decrease in afucosylated structure A2G2S(3,6)2 (T18). Taking the classification and clustering data into consideration Trf glycosylation does not provide superior differentiation between classes compared to the other glycoproteins- no clear PCA differentiation is observed

(Supporting Figure S10) but the hierarchical clustering does show its promise for resolving power between the groups (Supporting Figure S16).

With respect to clustering and classification analysis, the best performer is Hpt glycosylation (Figure 7) which displays clear groupings in the PCA analysis. In addition, regarding the statistical findings for Hpt glycosylation, there is a notable increase in two core fucosylated species FA2G2 (H11) and FA2G2S(6,6)2 (H20) in metastatic cancer patients (n=18) compared to normal controls (n=7) and a decrease in the afucosylated complex glycan A3G3S2 (H22). Taken collectively, this data suggests that core fucosylation (the FA2 Series) is significantly upregulated in this ovarian cancer cohort in both Trf and Hpt. Importantly, FA2 was previously found to be significantly altered (increased) in sera of ovarian cancer patients using an independent *N*-glycoanalytical technology[46]. We propose these changes reflect either differences in the expression levels of the α(1-6)-fucosyltransferase (FUT8) or donor substrate (GDP-fucose) in the *medial*-Golgi in ovarian cancer specific cells.

As discussed above, Trf glycosylation is the most significantly altered in our study and Hpt shows the greatest promise with respect to cluster analyses from the selected acute phase proteins and antibodies in the discrimination of ovarian cancer stages. Both glycoproteins warrant further functional studies in the future. Remarkably, very few investigations are described in the literature with respect to the possible role of Trf in ovarian cancer, despite its use as part of a commercial product, a multivariate index assay called OVA1 (approved by the FDA in 2016) which incorporates five serum biomarkers into a malignancy risk score of 0–10 using a proprietary algorithm and is recommended by the American College of Obstetricians and Gynecologists[48] as a tool to aid in evaluating women with adnexal masses. There is precedence of altered expression of Hpt glycosylation in ovarian cancer whereby Turner and colleagues found enhanced expression of branching, especially triantennary glycans in patients with ovarian cancer[49]. As such Trf and Hpt present as interesting targets for ovarian cancer treatment. Regarding the remaining glycoproteins, IgG, IgM and IgA glycosylation were assessed previously in the context of ovarian cancer from late stage patients only-the results indicated that combining IgG glycosylation profiles with CA-125 could improve accuracy of epithelial ovarian cancer prediction[18], in contrast to our findings. In a separate study, IgG galactosylation was used to assist a differential

diagnosis of ovarian cancer with CA-125[50]. A very early study investigated A1AT glycosylation in the context of ovarian cancer[51]. The authors concluded that fucosylation was a glycosylation feature that became elevated in the presence of tumour growth but remained low in remission and during chemotherapy. This study does not reflect the utility of A1AT fucosylation in the selected cohort of patients.

Conclusion:

This study presents the development of an elegant glycoanalytical platform for the detailed characterisation and investigation of a sequence of glycoprotein *N*-glycosylation: antibodies IgG, IgM and IgA and acute phase proteins Trf, Hpt and A1AT. In the course of this analysis, besides achieving a detailed glycoprofile for each glycoprotein through the identification and profiling of more than a hundred glycan structures, we were able to identify novel glycan motifs and traits that have not been outlined to date. Its utility in the context of ovarian cancer is highlighted-Trf and Hpt glycosylation can be exploited as biomarker tools for ovarian cancer and may play a cardinal role. More importantly, this comprehensive and reproducible glycoprofiling technology could provide significant insights and serve as a baseline for the identification of biomarkers and their regulation in a series of diseases.

Acknowledgments:
This work was supported by the EU FP7 programs High Glycan and GlycoBioM (278535, 259869). R.S. acknowledges funding from the Science Foundation Ireland Starting Investigator Research grant (SFI SIRG) under Grant Number 13/SIRG/2164. I.W. acknowledges funding from HighGlycoART: under Grant Number 1335G00086. The authors acknowledge the Leeds Multidisciplinary Research Tissue Bank for samples. The authors acknowledge Thermo Scientific for the donation of anti-haptoglobin capture resins.

Data Availability:
The mass spectrometry data in this study are freely available through the repository MassIVE (https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=a66ade995ac8431e80f6e27f11c55674, doi:10.25345/C5QD2X).

References:

1 Taniguchi, N. From glycobiology to systems glycobiology: international network with Japanese scientists through consortia. *IUBMB Life* **58**, 269-272, doi:10.1080/15216540600756020 (2006).
2 Butler, M. *et al.* Detailed glycan analysis of serum glycoproteins of patients with congenital disorders of glycosylation indicates the specific defective glycan processing step and provides an insight into pathogenesis. *Glycobiology* **13**, 601-622, doi:10.1093/glycob/cwg079 (2003).
3 Grunewald, S., Matthijs, G. & Jaeken, J. Congenital disorders of glycosylation: a review. *Pediatr Res* **52**, 618-624, doi:10.1203/00006450-200211000-00003 (2002).
4 Thanabalasingham, G. *et al.* Mutations in HNF1A result in marked alterations of plasma glycan profile. *Diabetes* **62**, 1329-1337, doi:10.2337/db12-0880 (2013).
5 Colhoun, H. O. *et al.* ANNALS EXPRESS: Validation of an automated UPLC IgG N-glycan analytical method applicable to Classical Galactosaemia. *Ann Clin Biochem*, 4563218762957, doi:10.1177/0004563218762957 (2018).
6 Freeze, H. H., Eklund, E. A., Ng, B. G. & Patterson, M. C. Neurological aspects of human glycosylation disorders. *Annu Rev Neurosci* **38**, 105-125, doi:10.1146/annurev-neuro-071714-034019 (2015).
7 Saldova, R. *et al.* Exploring the glycosylation of serum CA125. *Int J Mol Sci* **14**, 15636-15654, doi:10.3390/ijms140815636 (2013).
8 Sarrats, A. *et al.* Glycosylation of liver acute-phase proteins in pancreatic cancer and chronic pancreatitis. *Proteomics Clin Appl* **4**, 432-448, doi:10.1002/prca.200900150 (2010).
9 Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540-555, doi:10.1038/nrc3982 (2015).
10 Marino, K., Saldova, R., Adamczyk, B. & Rudd, P. M. Changes in Serum N-Glycosylation Profiles: Functional Significance and Potential for Diagnostics. *Spr Carb Ch* **37**, 57-93, doi:10.1039/9781849732765-00057 (2011).
11 Stockmann, H., Duke, R. M., Millan Martin, S. & Rudd, P. M. Ultrahigh throughput, ultrafiltration-based n-glycomics platform for ultraperformance liquid chromatography (ULTRA(3)). *Anal Chem* **87**, 8316-8322, doi:10.1021/acs.analchem.5b01463 (2015).
12 O'Flaherty, R., Trbojevic-Akmacic, I., Greville, G., Rudd, P. M. & Lauc, G. The sweet spot for biologics: recent advances in characterization of biotherapeutic glycoproteins. *Expert Rev Proteomics* **15**, 13-29, doi:10.1080/14789450.2018.1404907 (2018).

13 Lauc, G., Pezer, M., Rudan, I. & Campbell, H. Mechanisms of disease: The human N-glycome. *Bba-Gen Subjects* **1860**, 1574-1582, doi:10.1016/j.bbagen.2015.10.016 (2016).
14 McCarthy, C. *et al.* Increased outer arm and core fucose residues on the N-glycans of mutated alpha-1 antitrypsin protein from alpha-1 antitrypsin deficient individuals. *J Proteome Res* **13**, 596-605, doi:10.1021/pr400752t (2014).
15 Arnold, J. N. *et al.* Human serum IgM glycosylation: identification of glycoforms that can bind to mannan-binding lectin. *J Biol Chem* **280**, 29080-29087, doi:10.1074/jbc.M504528200 (2005).
16 Mattu, T. S. *et al.* The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fcalpha receptor interactions. *J Biol Chem* **273**, 2260-2272 (1998).
17 Bondt, A. *et al.* Longitudinal monitoring of immunoglobulin A glycosylation during pregnancy by simultaneous MALDI-FTICR-MS analysis of N- and O-glycopeptides. *Sci Rep* **6**, 27955, doi:10.1038/srep27955 (2016).
18 Ruhaak, L. R. *et al.* Protein-Specific Differential Glycosylation of Immunoglobulins in Serum of Ovarian Cancer Patients. *J Proteome Res* **15**, 1002-1010, doi:10.1021/acs.jproteome.5b01071 (2016).
19 Zhu, R., Zacharias, L., Wooding, K. M., Peng, W. & Mechref, Y. Glycoprotein Enrichment Analytical Techniques: Advantages and Disadvantages. *Methods Enzymol* **585**, 397-429, doi:10.1016/bs.mie.2016.11.009 (2017).
20 Stockmann, H., Adamczyk, B., Hayes, J. & Rudd, P. M. Automated, high-throughput IgG-antibody glycoprofiling platform. *Anal Chem* **85**, 8841-8849, doi:10.1021/ac402068r (2013).
21 Stockmann, H., O'Flaherty, R., Adamczyk, B., Saldova, R. & Rudd, P. M. Automated, high-throughput serum glycoprofiling platform. *Integr Biol (Camb)* **7**, 1026-1032, doi:10.1039/c5ib00130g (2015).
22 Henning Stoeckman, R. M. D. a. P. M. R. ULTRA3: Ultrafiltration-based, ultra-high throughput N-glycomics platform for ultra-performance liquid chromatography. (2015).
23 O'Flaherty, R. *et al.* Aminoquinoline Fluorescent Labels Obstruct Efficient Removal of N-Glycan Core alpha(1-6) Fucose by Bovine Kidney alpha-l-Fucosidase (BKF). *J Proteome Res* **16**, 4237-4243, doi:10.1021/acs.jproteome.7b00580 (2017).
24 Aitchison, J. *The statistical analysis of compositional data*. (Chapman \& Hall, Ltd., 1986).
25 Crookston, N. L. & Finley, A. O. yaImpute: An R package for kNN imputation. *J Stat Softw* **23** (2008).
26 Blashfield, R. K. Mixture Model Tests of Cluster-Analysis - Accuracy of 4 Agglomerative Hierarchical Methods. *Psychol Bull* **83**, 377-388, doi:Doi 10.1037//0033-2909.83.3.377 (1976).
27 Lauc, G., Vuckovic, F., Bondt, A., Pezer, M. & Wuhrer, M. Trace N-glycans including sulphated species may originate from various plasma glycoproteins and not necessarily IgG. *Nat Commun* **9**, 2916, doi:10.1038/s41467-018-05173-w (2018).
28 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).
29 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).
30 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).
31 Lauc, G. *et al.* Genomics Meets Glycomics-The First GWAS Study of Human N-Glycome Identifies HNF1 alpha as a Master Regulator of Plasma Protein Fucosylation. *Plos Genet* **6**, doi:ARTN e1001256

10.1371/journal.pgen.1001256 (2010).

32 Pucic, M. *et al.* High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol Cell Proteomics* **10**, M111 010090, doi:10.1074/mcp.M111.010090 (2011).
33 Zhao, S. *et al.* GlycoStore: a database of retention properties for glycan analysis. *Bioinformatics* **34**, 3231-3232, doi:10.1093/bioinformatics/bty319 (2018).
34 Royle, L. *et al.* Secretory IgA N- and O-glycans provide a link between the innate and adaptive immune systems. *J Biol Chem* **278**, 20140-20153, doi:10.1074/jbc.M301436200

M301436200 [pii] (2003).
35 Fujimura, T. *et al.* Glycosylation status of haptoglobin in sera of patients with prostate cancer vs. benign prostate disease or normal subjects. *Int J Cancer* **122**, 39-49, doi:10.1002/ijc.22958 (2008).
36 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).
37 Liu, S. D. *et al.* Afucosylated antibodies increase activation of FcgammaRIIIa-dependent signaling components to intensify processes promoting ADCC. *Cancer Immunol Res* **3**, 173-183, doi:10.1158/2326-6066.CIR-14-0125 (2015).
38 Pereira, N. A., Chan, K. F., Lin, P. C. & Song, Z. The "less-is-more" in therapeutic antibodies: Afucosylated anti-cancer antibodies with enhanced antibody-dependent cellular cytotoxicity. *MAbs* **10**, 693-711, doi:10.1080/19420862.2018.1466767 (2018).
39 Liang, J. X., Liang, Y. & Gao, W. Clinicopathological and prognostic significance of sialyl Lewis X overexpression in patients with cancer: a meta-analysis. *Onco Targets Ther* **9**, 3113-3125, doi:10.2147/OTT.S102389 (2016).
40 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* **57**, 289-300 (1995).
41 Rockfield, S., Raffel, J., Mehta, R., Rehman, N. & Nanjundan, M. Iron overload and altered iron metabolism in ovarian cancer. *Biol Chem* **398**, 995-1007, doi:10.1515/hsz-2016-0336 (2017).
42 Gornik, O. & Lauc, G. Glycosylation of serum proteins in inflammatory diseases. *Dis Markers* **25**, 267-278 (2008).
43 Sturiale, L. *et al.* Hypoglycosylation with increased fucosylation and branching of serum transferrin N-glycans in untreated galactosemia. *Glycobiology* **15**, 1268-1276, doi:10.1093/glycob/cwj021 (2005).
44 Zhang, S., Shang, S., Li, W., Qin, X. & Liu, Y. Insights on N-glycosylation of human haptoglobin and its association with cancers. *Glycobiology* **26**, 684-692, doi:10.1093/glycob/cww016 (2016).
45 Moss, E. L., Hollingworth, J. & Reynolds, T. M. The role of CA125 in clinical practice. *J Clin Pathol* **58**, 308-312, doi:10.1136/jcp.2004.018077 (2005).
46 Saldova, R. *et al.* Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiology* **17**, 1344-1356, doi:10.1093/glycob/cwm100 (2007).
47 Scholler, N. & Urban, N. CA125 in ovarian cancer. *Biomark Med* **1**, 513-523, doi:10.2217/17520363.1.4.513 (2007).
48 American College of, O. & Gynecologists' Committee on Practice, B.-G. Practice Bulletin No. 174: Evaluation and Management of Adnexal Masses. *Obstet Gynecol* **128**, e210-e226, doi:10.1097/AOG.0000000000001768 (2016).
49 Turner, G. A., Goodarzi, M. T. & Thompson, S. Glycosylation of alpha-1-proteinase inhibitor and haptoglobin in ovarian cancer: evidence for two different mechanisms. *Glycoconj J* **12**, 211-218 (1995).
50 Qian, Y. *et al.* Quantitative analysis of serum IgG galactosylation assists differential diagnosis of ovarian cancer. *J Proteome Res* **12**, 4046-4055, doi:10.1021/pr4003992 (2013).

51 Thompson, S., Guthrie, D. & Turner, G. A. Fucosylated Forms of Alpha-1-Antitrypsin That Predict Unresponsiveness to Chemotherapy in Ovarian-Cancer. *Brit J Cancer* **58**, 589-593, doi:DOI 10.1038/bjc.1988.265 (1988).

Figures:

Figure 1. Multiplexed automated serial capture of glycoprotein *N*-glycoprofiling. 96-well format robotic platform (A), specific anti-glycoprotein capture resin packed in PhyNexus phytip (B), serial capture of selected glycoproteins 1.Trf, 2.IgG, 3.IgM, 4.IgA, 5.Hpt, 6.A1AT (C), 1D SDS-PAGE separation of multiplexed automated capture of six selected glycoproteins from pooled human serum using PhyNexus Phytips. Lane 1: Protein Marker, Lane 2: Trf Standard, Lane 3: Bound Trf, Lane 4: Bound IgG, Lane 5: Bound IgM, Lane 6: Bound IgA, Lane 7: A1AT Standard, Lane 8: Bound A1AT, Lane 9: Hpt Standard, Lane 10: Bound Hpt. The protein bands marked with arrows are traces of albumin protein (non-glycosylated) (D), automated glycoprotein sample preparation, PNGaseF release and aminoquinoline carbamate (AQC) labelling of *N*-glycans (E) and finally ultra-high performance liquid chromatography (UPLC) separation and glycan structural analysis (F).

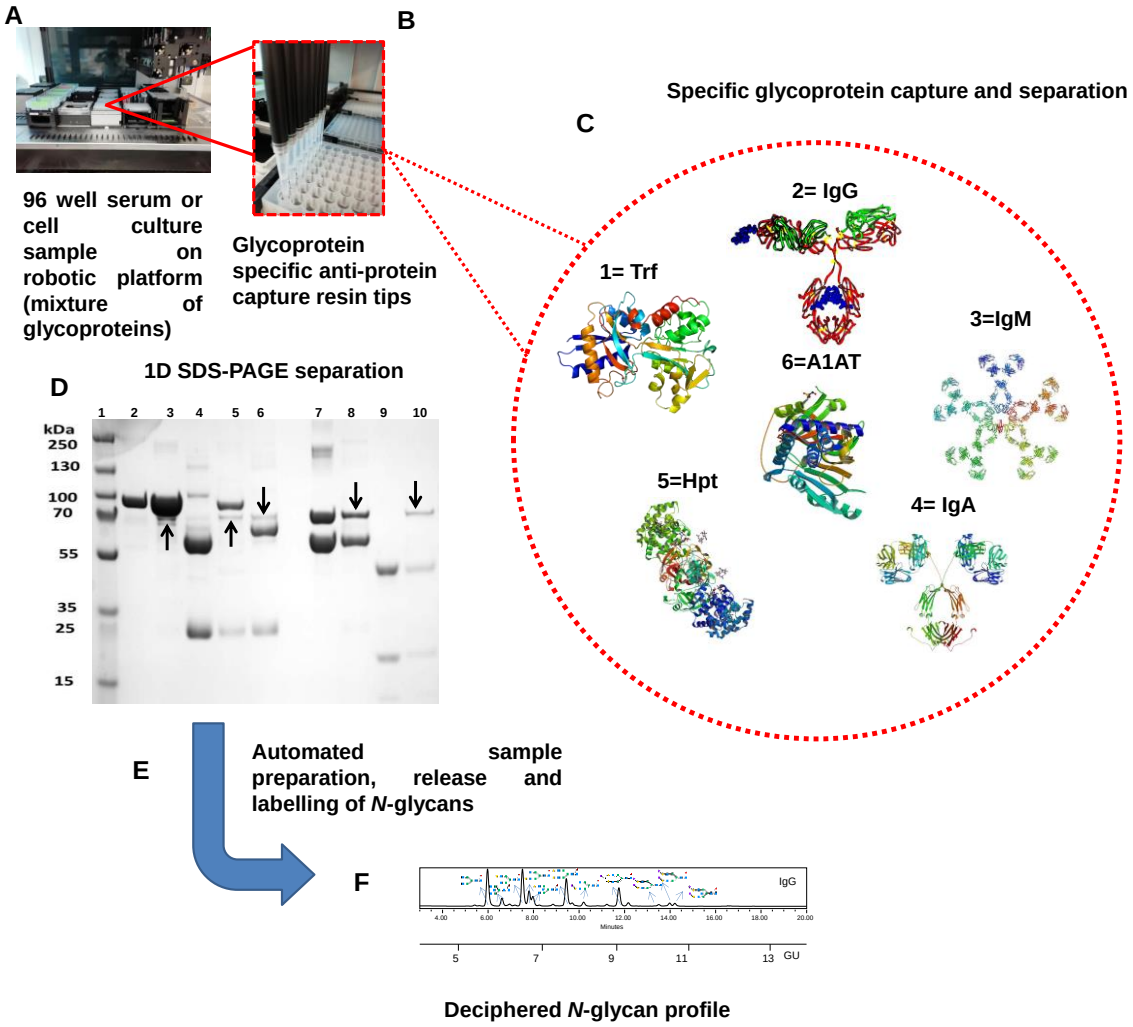


Figure 2. UPLC-HILIC-FLD chromatograms of AQC labelled *N*-glycans released from human serum for affinity purified glycoproteins IgG, IgM and IgA, Trf, Hpt, A1AT and 2-AB labelled *N*-glycans released from total serum. Only major glycans are annotated. All major *N*-glycans identified within the total serum UPLC chromatogram are present in the individual glycoprotein chromatograms which showcases that serum glycosylation is largely dominated by the highest abundance proteins (IgG, IgM and IgA, Trf, Hpt and A1AT).

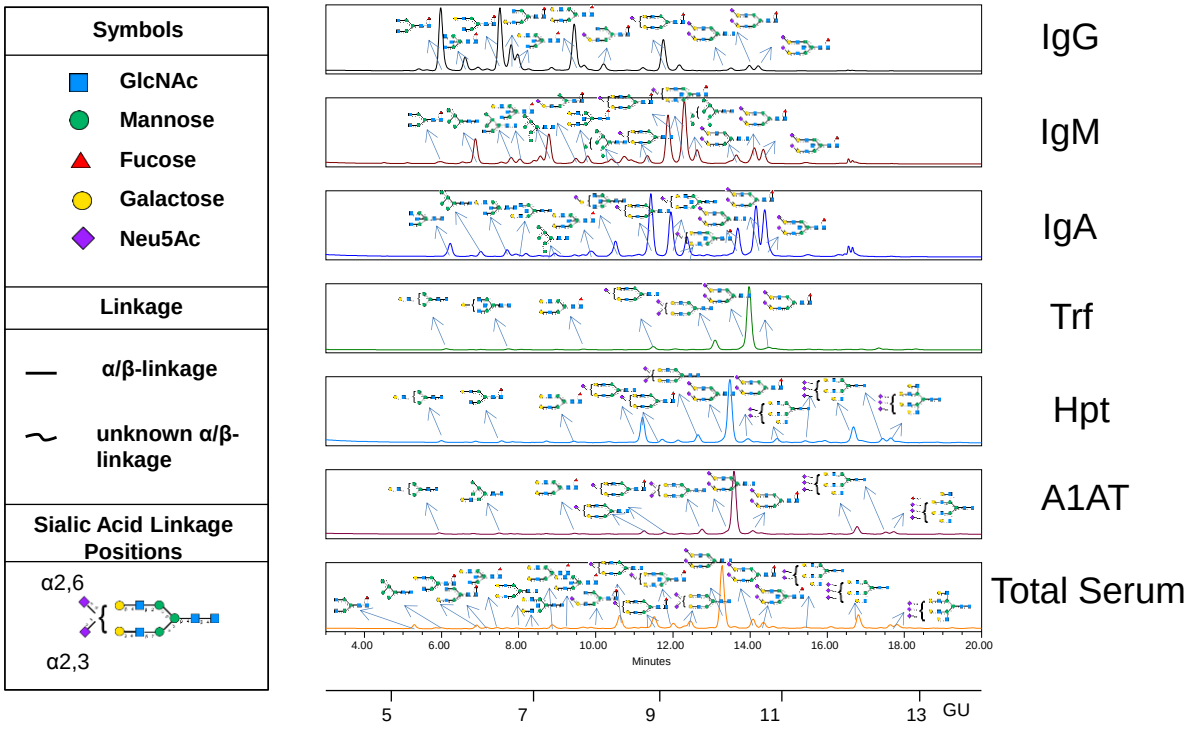


Figure 3. Serial capture and *N*-glycoprofiling of human serum IgG, IgM and IgA, purified from human serum. The 25 IgG glycan peak areas (G1-G25), 24 IgM peak areas (M1-24) and 25 IgA glycan peak areas (A1-A25) plotted for five technical replicates over three different days (3a). The standard error shown as error bars. Sequencing of AQC labelled IgG, IgM and IgA *N*-glycans visualized by UPLC-HILIC chromatograms using exoglycosidase enzymes with glucose units (GU) to facilitate glycan identification (3b). Digestion of AQC labelled IgG, IgM and IgA *N*-glycans with addition of sialidase (ABS), galactosidase (BTG), hexosaminidase (GUH), fucosidase (BKF) and mannosidase (JBM) in the following order ABS, ABS+BTG, ABS+BTG+GUH, ABS+BTG+GUH+BKF and ABS+BTG+GUH+BKF+JBM. For IgG, arrows indicate the cleavage of sugar residues for selected peaks: major glycans FA2G2S1 and FA2G2S2. For IgM, arrows indicate the cleavage of sugar residues for selected peaks: major glycans FA2G2S1 and FA2BG2S1. For IgA, arrows indicate the cleavage of sugar residues for selected peaks: major glycans A2G2S1, FA2G2S2 and FA2BG2S2. SNFG nomenclature is used for glycan representation.

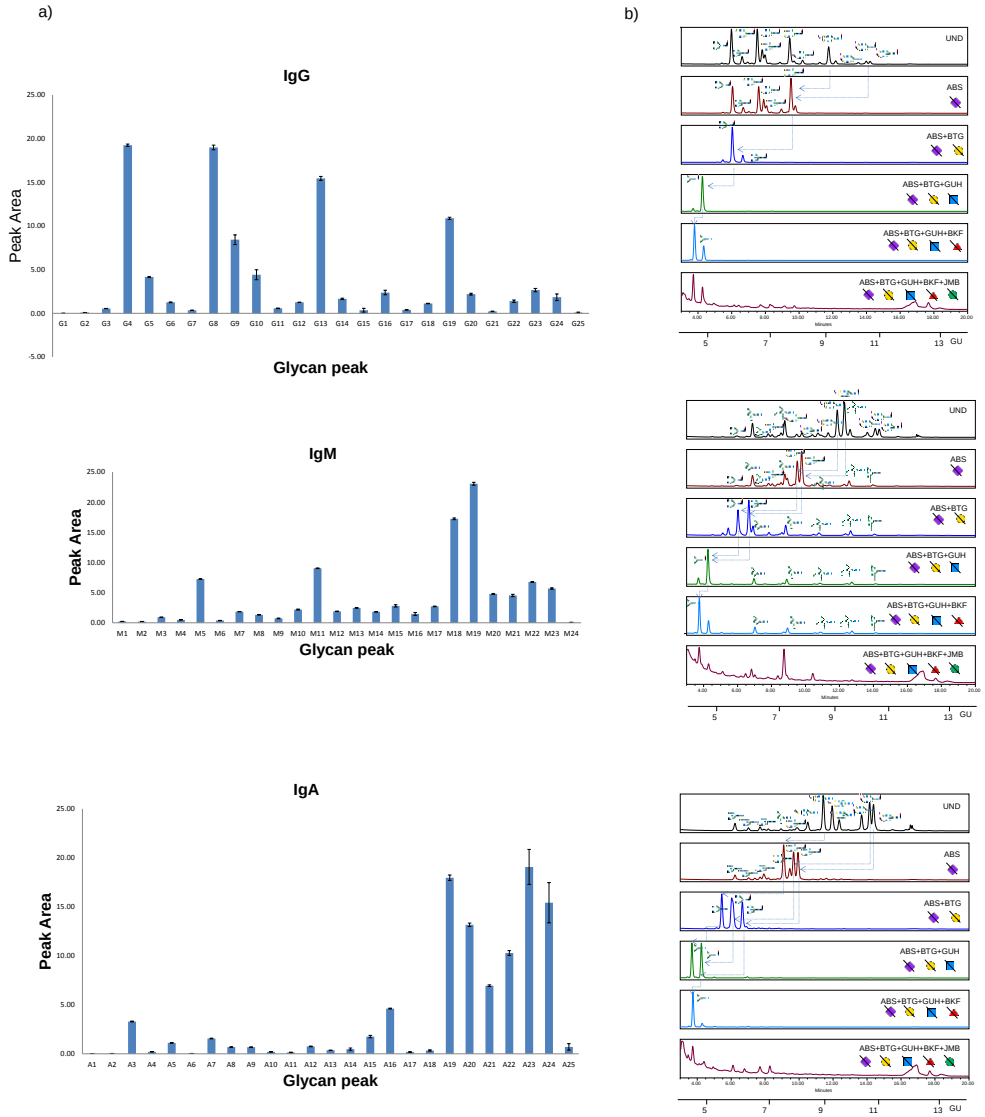


Figure 4. Serial capture and *N*-glycoprofiling of human serum Trf, Hpt and A1AT, purified from human serum. The 28 Trf glycan peak areas (T1-T28), 31 Hpt glycan peak areas (H1-H31) and 28 A1AT glycan peak areas (AT1-AT28) were plotted for five technical replicates over three different days (4a). The standard error shown as error bars. Sequencing of AQC labelled Trf, Hpt and A1AT *N*-glycans visualized by UPLC-HILIC chromatograms using exoglycosidase enzymes with glucose units (GU) to facilitate glycan identification (4b). Digestion of AQC labelled Trf, Hpt and A1AT *N*-glycans with addition of sialidase (ABS), galactosidase (BTG), hexosaminidase (GUH), fucosidase (BKF) and mannosidase (JBM) in the following order ABS, ABS+BTG, ABS+BTG+GUH, ABS+BTG+GUH+BKF and ABS+BTG+GUH+BKF+JBM. In Trf, arrows indicate the cleavage of sugar residues for selected peaks: major glycans A2G2S(3,6)2 and A2G2S2(6,6)2. In Hpt, arrows indicate the cleavage of sugar residues for selected peaks: major glycans A3G3S3, A2G2S2 and A2G2S1. In A1AT, arrows indicate the cleavage of sugar residues for selected peaks: major glycans A3G3S3 and A2G2S2.SNFG nomenclature is used for glycan representation.

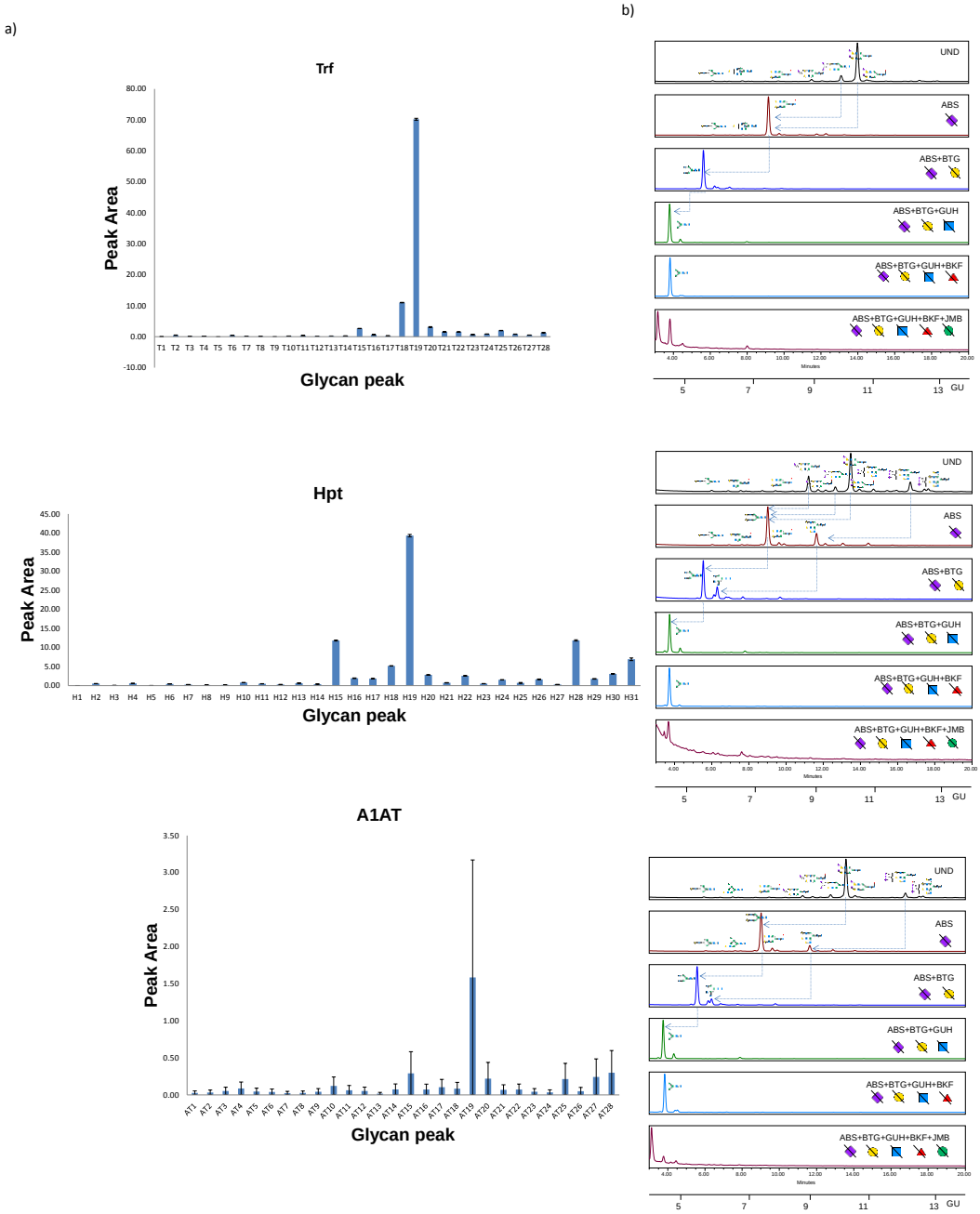
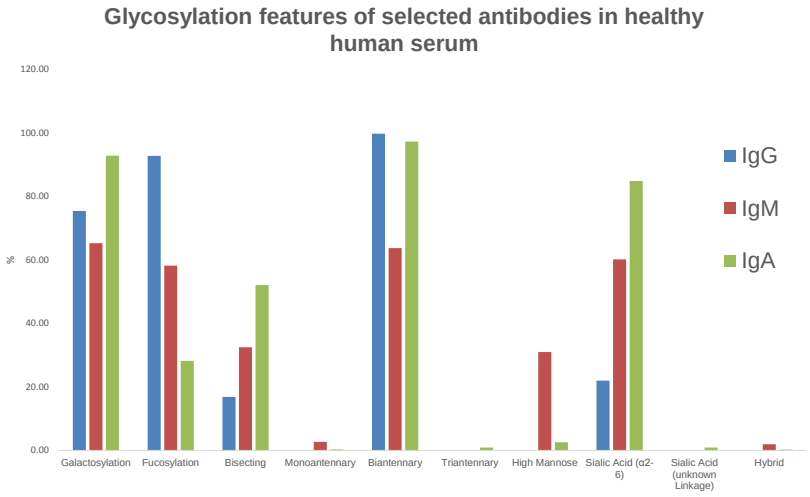


Figure 5. Glycosylation traits for selected antibodies (IgG, IgM, IgA) and acute phase proteins (Trf, Hpt, A1AT) for healthy human serum are presented (5a). The values are represented as a % of the total glycosylation and are calculated from data provided in Supporting Table S15. Glycosylation features of antibodies (IgG, IgM and IgA) and acute phase proteins (Trf, Hpt, A1AT) are presented as control charts (5b and 5c).

a)

Glycan Traits	IgG	IgM	IgA	Trf	Hpt	A1AT
Galactosylation	75.44	65.27	92.87	98.31	96.87	97.77
Fucosylation	92.80	58.21	28.22	8.07	16.16	14.48
Bisecting	16.87	32.48	52.09	1.13	0.31	3.44
Monoantennary	0.09	2.72	0.31	0.97	0.96	1.09
Biantennary	99.78	63.75	97.28	89.94	67.93	80.62
Triantennary	0.00	0.00	0.90	7.04	22.45	12.14
High Mannose	0.00	30.98	2.58	1.01	2.77	1.92
SLe_x Motit	0.00	0.00	0.00	1.56	1.48	0.65
Sialic Acid (α2-3)	0.00	0.00	0.00	11.57	6.24	5.19
Sialic Acid (α2-6)	21.99	60.20	84.87	71.85	65.43	78.99
Sialic Acid (unknown Linkage)	0.00	0.00	0.90	8.13	27.16	15.13
Hybrid	0.00	1.94	0.29	0.38	0.00	0.97
Tetraantennary	0.00	0.00	0.00	0.00	5.59	0.00

b)



c)

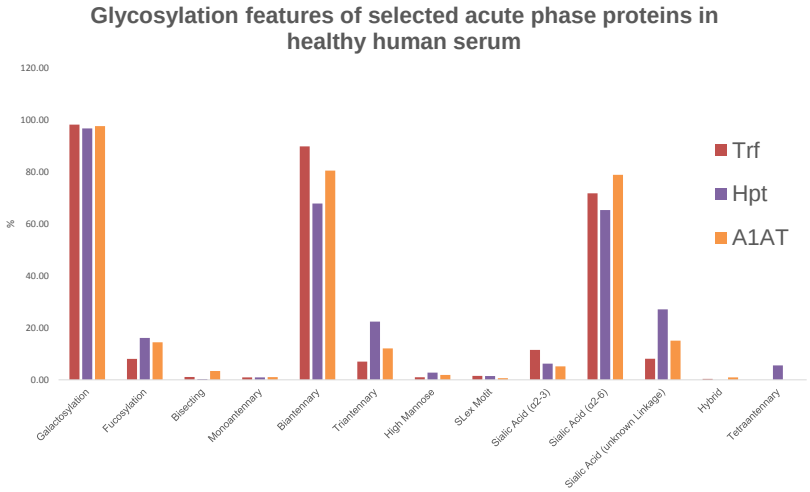
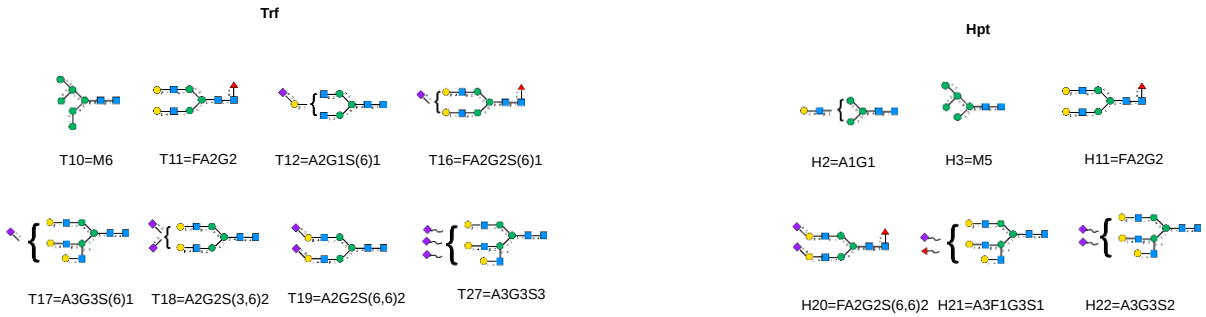


Figure 6. Statistically significant GPs and glycosylation traits for each glycoproteins and clinical parameters (Y=yes) with p-values <0.05 (corrected for multiple testing error using a 5% FDR proposed by Benjamini-Hochberg[40]) for ovarian cancer cohort of normal vs. borderline, metastatic vs. borderline and normal vs. metastatic (6a). The major glycan for the each statistically significant GP is presented (6b). Boxplots are presented for the statistically significant GPs, glycosylation traits and logCA125 for borderline (red), metastatic (green) and normal (blue) samples (6c).

a)

	Normal vs. Borderline	Metastatic vs. Borderline	Normal vs. Metastatic
IgG	-	-	-
IgM	-	-	-
IgA	-	-	-
Trf	T11, T16, T17, T18, fucosylation	-	T10, T11, T12, T16, T17, T18, T19, T27, fucosylation, biantennary, sialic acid (α2,6)
Hpt	-	-	H2, H3, H11, H20, H21, H22, fucosylation, SLex
A1AT	-	-	-
Age	-	-	-
CA125	-	Y	Y
CRP	-	-	-
Protein	-	-	-
Menopause	-	-	-

b)



c)

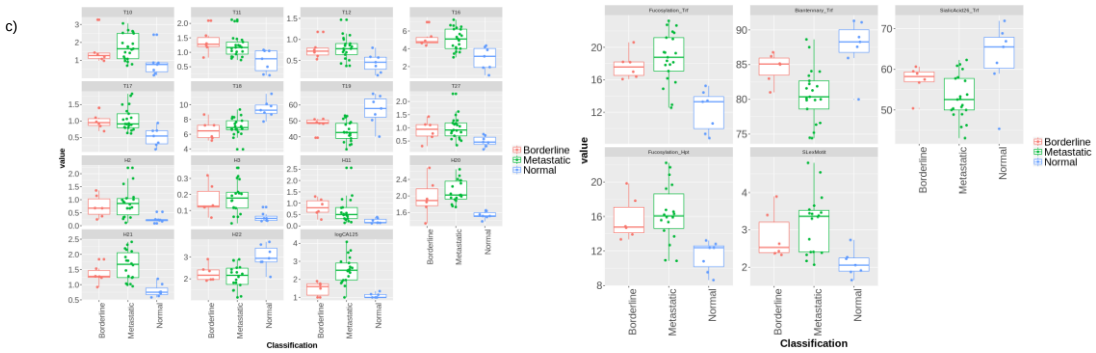


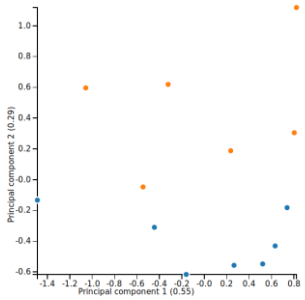
Figure 7. Discrimination performance for noteworthy glycosylation traits (Hpt) for ovarian cancer cohort of normal vs. borderline, normal vs. metastatic and borderline vs metastatic. Linear regression model including AUC, SEN and SPE (7a), cluster analysis using PCA using Hpt derived traits as input for normal (blue, n=7) vs. borderline (orange, n=6) separation, normal (blue, n=7) vs. metastatic (orange, n=18) and borderline (orange, n=6) vs metastatic blue, n=18) respectively. In brackets the variance of the principal component.

a)

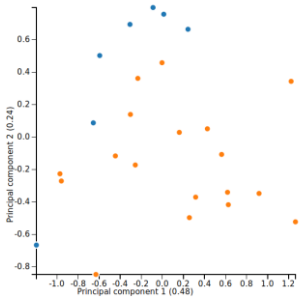
	Log CA125	All peaks	Derived traits
Hpt	AUC:0.667	AUC:0.976	**AUC:1.000**
Normal (n=7)	SEN:0.667	SEN:0.833	**SEN:1.000**
Borderline (n=6)	SPE:1.000	SPE:1.000	**SPE:1.000**
Hpt	**AUC:1.000**	**AUC:1.000**	**AUC:1.000**
Normal (n=7)	**SEN:1.000**	**SEN:1.000**	**SEN:1.000**
Metastatic (n=18)	**SPE:1.000**	**SPE:1.000**	**SPE:1.000**
Hpt	**AUC:0.917**	AUC:0.440	AUC:0.694
Borderline (n=6)	**SEN:0.889**	SEN:0.667	SEN:0.556
Metastatic (n=18)	**SPE: 1.000**	SPE:0.500	SPE:0.833

b)

Hpt derived traits
Normal vs. borderline



Hpt derived traits
Normal vs. metastatic



Hpt derived traits
Borderline vs. metastatic