Glycan biomarkers of autoimmunity and bile acid-associated alterations of the human glycome: primary biliary cirrhosis and primary sclerosing cholangitis-specific glycans

Maverakis, E.; Merleev, A.A.; Park, D.; Kailemia, M.J.; Xu, G.G.; Ruhaak, L.R.; ... ; Lebrilla, C.B.

Clinical Immunology 230 (2021) 108825

Contents lists available at ScienceDirect

Clinical Immunology

journal homepage: www.elsevier.com/locate/yclim

Full Length Article

Glycan biomarkers of autoimmunity and bile acid-associated alterations of the human glycome: Primary biliary cirrhosis and primary sclerosing cholangitis-specific glycans

Emanual Maverakis [a,*], Alexander A. Merleev [a], Dayoung Park [b,c,d], Muchena J. Kailemia [b], Gege Xu [b], L. Renee Ruhaak [b,e], Kyoungmi Kim [f], Qiuting Hong [b], Qiongyu Li [b], Patrick Leung [g], William Liakos [a], Yu-Jui Yvonne Wan [h], Christopher L. Bowlus [i], Alina I. Marusina [a], Nelvish N. Lal [a], Yixuan Xie [b], Guillaume Luxardi [a], Carlito B. Lebrilla [b,j,k]

[a] Department of Dermatology, University of California Davis School of Medicine, Sacramento, CA, USA
[b] Department of Chemistry, University of California Davis, Davis, CA, USA
[c] Department of Surgery, Center for Drug Discovery and Translational Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
[d] Wyss Institute of Biologically Inspired Engineering, Harvard University, Boston, MA, USA
[e] Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, ZA, Leiden, the Netherlands
[f] Division of Biostatistics, Department of Public Health Sciences, University of California Davis, Davis, CA, USA
[g] Department of Internal Medicine, Division of Rheumatology, Allergy and Clinical Immunology, University of California Davis School of Medicine, Davis, CA, USA
[h] Department of Medical Pathology and Laboratory Medicine, University of California Davis School of Medicine, Sacramento, CA, USA
[i] Division of Gastroenterology and Hepatology, UC Davis School of Medicine, CA, USA
[j] Department of Biochemistry and Molecular Medicine, University of California Davis, Davis, CA, USA
[k] Foods for Health Institute, University of California Davis, Davis, CA, USA

ARTICLE INFO

Keywords:
N-glycopeptides
Multiple Reaction Monitoring
plasma
biomarker testing
Primary Biliary Cirrhosis (PBC)
Primary Sclerosing Cholangitis (PSC)

ABSTRACT

We have recently introduced multiple reaction monitoring (MRM) mass spectrometry as a novel tool for glycan biomarker research and discovery. Herein, we employ this technique to characterize the site-specific glycan alterations associated with primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC). Glycopeptides associated with disease severity were also identified. Multinomial regression modelling was employed to construct and validate multi-analyte diagnostic models capable of accurately distinguishing PBC, PSC, and healthy controls from one another (AUC = 0.93 ± 0.03). Finally, to investigate how disease-relevant environmental factors can influence glycosylation, we characterized the ability of bile acids known to be differentially expressed in PBC to alter glycosylation. We hypothesize that this could be a mechanism by which altered self-antigens are generated and become targets for immune attack. This work demonstrates the utility of the MRM method to identify diagnostic site-specific glycan classifiers capable of distinguishing even related autoimmune diseases from one another.

1. Introduction

Glycans (i.e. oligosaccharides or sugars) are one of the four fundamental molecules that make up all living systems [1]. The totality of glycans within an organism is the glycome. The process that synthesizes and enzymatically attaches glycans to organic molecules is called glycosylation [2]. Cell surface and extracellular proteins are commonly post-translationally modified with glycans, which can fine-tune protein function by acting as "on and off" switches or as "analog regulators" [3]. However because there is no template for glycan synthesis, it is extremely difficult to predict the composition of the human glycome from gene expression data. In fact, when one considers the massive 3-

Abbreviations: MRM, multiple reaction monitoring; primary biliary cirrhosis, PBC; primary sclerosing cholangitis, PSC; CID, collision induced dissociation; VIF, variance inflation factor; CDCA, chenodeoxycholic acid; OCA, obeticholic acid; GCDCA, glycochenodeoxycholate; LCA, lithocholic acid; DCA, deoxycholic acid; ConA, Concanavalin A; JAC, Jacalin from *Artocarpus integrifolia*; SNA, *Sambucus nigra* agglutinin; AAL, *Aleuria aurantia* lectin.

* Corresponding author at: UC Davis Medical Center, Department of Dermatology, 3301 C Street Suite 1400, Sacramento, CA 95816, USA.
E-mail address: emaverakis@ucdavis.edu (E. Maverakis).

https://doi.org/10.1016/j.clim.2021.108825
Received 27 January 2021; Received in revised form 10 August 2021; Accepted 11 August 2021
Available online 14 August 2021

dimensional structural diversity of glycans combined with their variation in attachment sites, the complexity of the glycome parallels that of the genome [3].

The National Research Council of the U.S. National Academies has highlighted the importance of glycans as biomarkers of human disease [3]. It is believed that glycans play a key role in the pathophysiology of all human diseases. With respect to autoimmune disease, we put forth the Altered Glycan Theory of Autoimmunity, which argues that each autoimmune disease will have a unique glycan signature [4]. Most prior glycan biomarker studies in autoimmunity have used labor-intensive methodologies to characterize glycans released from purified proteins and perhaps for this reason, detailed analyses have only been conducted on a relatively small number of patients. Lower resolution techniques, which yield limited structural information and no site-specific information, have been used to characterize larger patient cohorts, but such untargeted analyses are not ideally suited for biomarker discovery and reproducibility. As a result, multi-analyte glycan biomarkers capable of distinguishing one autoimmune disease from another have not yet been developed.

With the goal of deploying glycan biomarkers clinically, we have developed Multiple Reaction Monitoring (MRM) to characterize the human plasma glycome in a rapid, site-specific, and reproducible fashion [5]. Although MRM mass spectrometry (MS) is mainly used in the fields of metabolomics and proteomics [6–9], its high sensitivity and linear response over a wide dynamic range makes it especially suited for glycan detection [10]. Herein, we employ MRM MS to characterize the glycan alterations associated with PBC (primary biliary cirrhosis; an autoimmune disease of the intrahepatic bile ducts) and PSC (primary sclerosing cholangitis; an autoimmune disease of the intrahepatic and extrahepatic bile ducts). Using multinomial logistic regression we construct a multi-analyte classifier model capable of distinguishing PBC, PSC, and healthy controls from one another. Finally, as a demonstration that disease-relevant environmental factors can alter glycosylation, the ability of bile acids to alter B-cell glycosylation was assessed. This study represents a viable alternative to existing diagnostic technology tools that has the potential to define specific glycan-based markers of autoimmune disease with minimal patient sample and reliable quantification.

2. Material and Methods

2.1. Study design

The objective of this study was to identify the relative abundance of site-specific glycosylations within the most abundant plasma proteins and then to use this information to distinguish multi-analyte classifiers capable of differentiating autoimmune liver diseases. PBC and PSC patients were recruited by the University of California (UC) Davis, Department of Medicine, Division of Gastroenterology and Hepatology. All patients met the clinical diagnostic criteria of PBC or PSC as defined by the American Association for the Study of Liver Diseases (AASLD) [11,12]. As per the AASLD guidelines a biopsy diagnosis was not required if other diagnostic criteria were met. There were no established upper or lower age limits for enrollment. Healthy individuals were recruited from the UC Davis Medical Center as controls. The UC Davis institutional review board approved this study. All participants provided their written informed consent. Demographics are listed in Table S1.

2.2. Sample preparation

For each individual enrolled, plasma was separated from whole blood using a Ficoll gradient. From each plasma preparation a 2 μL aliquot was reduced, alkylated and then subjected to trypsin digestion at 37 °C as previously described [13]. To allow for absolute protein quantification, 100 μg of IgG, IgA, and IgM (all from Sigma-Aldrich, St. Louis, MO) was digested according to the same protocol and a dilution series was made prior to sample injection.

2.3. UPLC-ESI-QqQ-MS analysis

The neat enzymatically-prepared samples containing both peptides and glycopeptides were then directly analyzed without further hands-on sample cleanup or dilution using an Agilent 1290 infinity liquid chromatography (LC) system coupled to an Agilent 6490 triple quadrupole (QqQ) mass spectrometer (Agilent Technologies, Santa Clara, CA), as previously described [13,14]. Briefly, an Agilent Eclipse plus C18 precolumn (RRHD 1.8 μm, 2.1 X 5 mm) was connected directly to an Agilent Eclipse plus C18 column (RRHD 1.8 μm, 2.1 X 100 mm) which was used for UPLC separation. 1.0 μL of the digested plasma samples was injected and analyzed using a 25-min binary gradient consisting of solvent A of 3% acetonitrile, 0.1% formic acid and solvent B of 90% acetonitrile, 0.1% formic acid in nano-pure water (*v*/v) at a flow rate of 0.5 mL/min.

The MRM method used for this study requires predetermined knowledge of the peptide or glycoforms' LC retention time, its electrospray ionization, and its collision induced dissociation (CID) behavior, which we have previously determined for all the non-glycosylated peptides and glycopeptides used in this study [5,13]. Results were integrated using Agilent MassHunter Quantitative Analysis B.5.0 software. Protein concentrations were determined based on calibration curves and glycopeptide relative responses were calculated using the area under the curves of the glycopeptide and a non-glycosylated reference peptide from the same protein. Relative IgM is IgM adjusted for high affinity IgG.

2.4. Bile acid-induced glycan alterations

2.4.1. Raji cell culture, bile acid treatment, and flow cytometry

For this analysis, Raji B cells were *in vitro* cultured in triplicate with different dilutions of the following bile acids representing both disease-relevant and physiologic concentrations: chenodeoxycholic acid (CDCA), obeticholic acid (OCA), glycochenodeoxycholate (GCDCA), lithocholic acid (LCA), and deoxycholic acid (DCA). After incubation with the different bile acids, Raji B cells were incubated with Fc-block (BD Biosciences, San Jose, CA) on ice for 15 min and stained with Aqua-LIVE/DEAD (Invitrogen, Carlsbad, CA). Cells were then stained with or without (FMO, Fluorescence Minus One) Fluorescein labeled lectins Concanavalin A (ConA), Jacalin from *Artocarpus integrifolia* (JAC), *Sambucus nigra* agglutinin (SNA), and *Aleuria aurantia* lectin (AAL) (Vector Laboratories, Burlingame, CA) for 30 min at room temperature. Cells were washed after each step and before being analyzed on a BD Fortessa flow cytometer (BD, Franklin Lakes, NJ). Data were analyzed using FlowJo (Tree Star, Ashland, OR). After gating single cells and live cells, Fluorescein labeled lectins geometric mean fluorescence was retrieved and plotted in a heat map using R (R Core Team, Vienna, Austria) to represent bile-acid induced alterations of the B-cell glycocalyx. Standard mean differences (SMD) were calculated from mean fluorescence intensities of the bile acid-culture B cells and normalized to control cultured cells.

2.4.2. RNA isolation and qRT-PCR

Raji B cells were harvested, washed twice with PBS, and resuspended in RNAlater (Life Technologies, Carlsbad, CA). Total RNA were extracted using RNeasy plus mini kit (Qiagen, Germantown, MD) and the quantity and quality of RNA were determined by using a Qubit Fluorometer (Life Technologies) and TapeStation 2200 (Agilent Technologies, Santa Clara, CA) following the manufacturer's protocol. Total RNA was reverse transcribed to cDNA using iScript Reverse Transcription Supermix (Bio-RadLaboratories, Hercules, CA) following the manufacturer's protocol. Predesigned human glycosylation and extracellular matrix and cytoskeleton PrimePCR plates (Bio-RadLaboratories) were used for real-time PCR using the CFX96 Touch Real-Time PCR detection system (Bio-RadLaboratories) and the analysis was performed using CFX

Manager 3.1 (Bio-RadLaboratories). Gene expression was normalized to reference genes and presented as fold changes.

2.5. Statistical analysis

All statistical analyses were done using R software [15]. For each analyte skewedness was assessed and data was log transformed to make the distribution approximately normal. Outliers were identified using R package "extreamvalues" [16], and when present were winsorized from the analysis so that the outliers were set equal to the nearest non-outlier value. Differential analysis was carried out to identify analytes whose concentration was significantly different between disease groups (PBC vs. PSC vs. control) using Analysis of Covariance (ANCOVA) with age and gender as covariates, followed by Tukey's honest significant difference (HSD) tests. ANCOVA and linear regression assumptions about the normality of residuals were examined by use of the Shapiro-Wilk test. Collinearity of variables in the multivariate models was examined by calculating variance inflation factor (VIF, excessive if >2.5) with R package "car" [17]. Nonlinear relationships between the analytes and the outcome were evaluated with R package "mfp" using a multiple fractional polynomial (MFP) method [18]. Variable selection in the multiple linear regressions analyses was performed by forward stepwise exhaustive search using "leaps" R package [19]. The algorithm searched the best models of all sizes up to the specified maximum number variables. Each model's performance was evaluated by the leave-one-out cross validation method using "caret" [20] R package and the optimal number of variables included in the model was selected using minimum root-mean-square error (RMSE). Logistic regression models were fitted using Firth's bias reduction method with the R package "logistf" [21]. This package also used penalized likelihood ratio tests for variable selection. The association between analytes and disease status (PBC, PSC, and control) was analyzed using multinomial logistic regression models with R package "nnet" [22]. Model performance was estimated with 5-fold cross validated (CV) multiple class area under ROC curve using R package "HandTill2001" [23]. The association between analytes and disease stages was examined using an ordinal logistic regression implemented in the R package "MASS" [22]. Proportional odds assumption was checked by Brant test for this regression in the R package "brant" [24]. Meta-analyses were conducted to assess findings across the multiple datasets using R package "metafor" [25]. A weighted random-effects model was used to estimate a summary effect size. Restricted maximum-likelihood estimator was selected to estimate between-study variance. Weighted estimation with inverse-variance weights was used to fit the model.

3. Results

3.1. Significant difference of glycoforms among PBC, PSC, and healthy controls

As a demonstration of the utility of the plasma glycome to identify biomarkers of human disease, accounting for age and gender as covariates, ANCOVA and Tukey's HSD test were performed to identify analytes that were significantly differentially expressed among three groups (PBC, PSC, and healthy control). Group demographics are summarized in Table S1 and Fig. S1. PBC patients differed significantly from controls with respect to 61 glycoforms (66 total analytes). PSC patients significantly differed from controls with respect to 56 glycoforms (60 total analytes) and there were 47 glycoforms (54 total analytes) that differed significantly between PBC and PSC (Fig. 1 and Table S2). This analysis also revealed that the relative abundance of IgG3 and IgM were significantly elevated (P = 6.43e-14 and 3.61e-14, respectively) in patients with PBC when compared to controls (0.1 ± 0.1 vs 0.041 ± 0.02 and 0.29 ± 0.2 versus 0.086 ± 0.07, respectively) (Table S2).

3.2. Altered glycosylation at different stages of PBC

Given that there were several glycoforms significantly differentially expressed in patients with PBC, it was of interest to determine if disease stage was also associated with alterations in glycosylation. A number of patients (15) lacking stage information were dropped from the analysis leaving 46 patients: 23 with Stage I disease, 13 with Stage II, 5 with Stage III and 5 with Stage IV disease (Table S1). For this analysis, glycoforms that were differentially expressed in the setting of PBC were graphed against stage, and Spearman's correlation coefficients (r_s) and corresponding P values for significance were calculated. Glycoforms

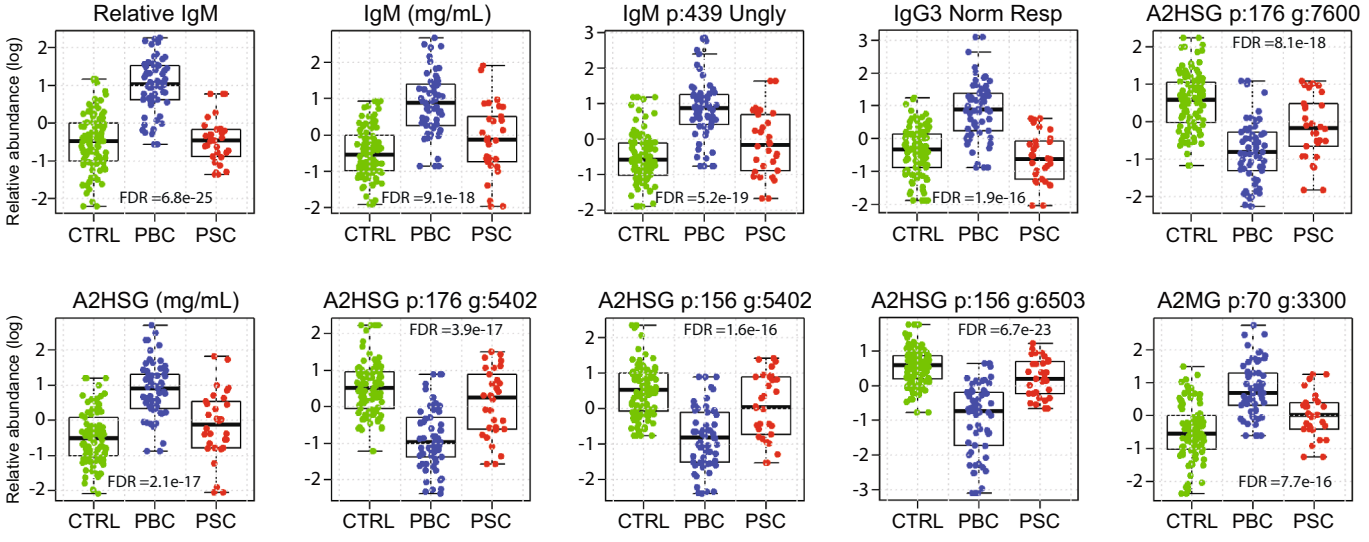


Fig. 1. Analysis of Covariance (ANCOVA) of glycan or protein relative abundances in PBC, PSC, and CTRL groups. Y axes represent the log of the relative abundance. Examples of the significant analytes that were differently expressed between PBC, PSC, and control groups (CTRL) are depicted. A full list of all monitored analytes including P values, relative concentrations, and standard deviations can be found in Table S2. The upper and lower bars connected to each box indicate the boundaries of the normal distribution and the upper and lower box edges mark the first and third quartile boundaries within each distribution. The bold line within the box indicates the median value of the distribution. FDR values are displayed for each analyte with analytes organized in order of increasing value going across each row left to right. Glycan nomenclature is as follows: p: site on protein, g: glycan code. Glycan conversion from code (i.e. g:3300) to chemical composition and structures (i.e.Hex(3)HexNAc(3)) can be seen in Fig. S3.

IgG2 3510, AGP1 93–6503, and AGP1 93–8704 (see Fig. S3 for nomenclature) were found to correlate with disease stage ($r_s = 0.5$, $P =$ 4e-04; $r_s = -0.43$, $P = 0.0028$; and $r_s = -0.43$, $P = 0.003$, respectively) (Fig. 2 and Table S3). For these patients, the effect of disease stage on glycosylation was also evaluated using ANOVA, which could verify the significant relationship of glycoforms AGP1 93–6503 and AGP1 93–8704 (Fig. S3) with disease stage (FDR = 3.56e-03 and 3.32e-02, respectively); however, IgG2 3510 response according to stage did not reach significance by ANOVA (Table S4). To account for the possibility that the increase in disease severity was nonuniform between incremental increases in disease stage, ordinal logistic regression was employed. This analysis demonstrated that the relative abundances of AGP1 93–6503, AGP1 93–8704, IgA1/2 p:144 g:4500, and IgG2 3510 were dependent on disease stage when stages I and II were merged in order to ensure proportionality of data ($P < 0.05$) (Table S5).

3.3. Differential expression of glycoforms in PBC, PSC, and healthy controls: multi-analyte classifier performance

Fig. 3A presents receiver operating characteristic (ROC) curves and area under the curve (AUCs) of differentially expressed glycoforms that performed best as single analyte multinomial classifiers capable of distinguishing between PBC, PSC, and control groups. Specifically, seven glycoforms and three plasma proteins (relative IgM, normalized IgG3, and absolute A2HSG) performed well (AUCs >0.72) performing as single analyte classifiers (Fig. 3A, Fig. S4A, and Table S6).

Multi-analyte classifiers were then constructed from analytes which demonstrated low Pearson's Product-Moment Correlation Coefficient (PPMCC) r values in their pairwise comparisons ($|r| < 0.2$). For each classification model, ROC curves were constructed and AUCs calculated to illustrate the classifier's diagnostic ability. Pairwise classifiers designed to distinguish PBC from controls, PSC from controls, and PBC from PSC performed their designated tasks well (5-fold CV AUCs = 0.98 ± 0.03; 0.96 ± 0.06; 0.94 ± 0.06; respectively) (Fig. 3B to D and Table S7). Final multinomial models (capable of differentiating all diagnostic groups from each other) with differing numbers of analytes ($n = 1$–13) were then assessed for accuracy. As additional analytes were added to the multinomial model, meaningful increases in accuracy were noted until four analytes were reached, above which only small additional increases in accuracy were observed (Fig. S4B). Thus, four analytes were chosen for the final multinomial model: Hp 207–11,904, A1AT 70–5402, Hp 241–5511 and relative IgM (Fig. 3E, Table S8). The AUC of this four-analyte multinomial model (capable of differentiating all diagnostic groups from each other) was 0.93 ± 0.03, 5-fold CV (Fig. 3E). Alternative models are presented in Fig. 3E and Fig. S4 for comparison. For all models described above, collinearity among analytes was evaluated by calculating their variance inflation factor (VIF) and found to be low (Table S7). Nonlinear relationships were evaluated by the MFP method, which revealed no concerning nonlinearity. Lastly, all constructed models were validated using the 5-fold CV technique.

3.4. Environmental influences on glycosylation

It has been stated that the plasma glycome is an expression of the overall health of an individual [4]. This is in part due to the multitude of environmental factors that impact glycosylation. Thus, we sought to identify environmental influences that could alter glycosylation in the setting of PBC and PSC. Given that PBC and PSC have differential elevation of bile acids compared to each other as well as healthy controls [26,27], we characterized the influence of bile acids on the expression of glycogenes in Raji B cells (Fig. 4A and Fig. S5). qPCR of the Raji B cells under different culture conditions demonstrated that some glycogenes (e.g. *GALNT1* and *GALNT9*) were consistently altered in response to co-culture with bile acids (Fig. 4A and Fig. S5). Reduced expression of *GALNT9*, which encodes a GalNAc transferase responsible for the initial transfer of a GalNAc residue to a serine or threonine residue on nascent proteins, in all bile acid conditions suggest that bile acids may shape the O-glycosylation pattern of B cells. Altered expression of MAN1C1, encoding a mannosidase that functions to glycosylate mature Asn-linked oligosaccharides, also appeared in the majority of bile acid conditions potentially indicating that N-glycosylation of B cells is also affected. The majority of glycogenes decreased in expression with LCA addition, including *GALNT1*, *GALNT9*, and *GALNT12*. Addition of GCDCA (5 μM) and OCA (0.5 μM) resulted in the least pronounced changes in glycogene expression while addition of CDCA (0.3 μM) and LCA (0.03 μM) resulted in the most varied transcripts among all bile acid conditions (Fig. 4A and Fig. S5).

In addition to changes in gene expression, we also sought to characterize the bile-acid induced alterations to the B cell glycocalyx. Thus, the same bile acid-B cell cultures described above were also analyzed by flow cytometry using an array of fluorescein labeled lectins, which are glycan-binding proteins with structural specificity. These experiments demonstrated that, when compared to bile acids LCA and CDCA, the bile acids OCA, DCA, and GCDCA produced opposing glycan alterations in Raji B cells. After exposure to the second group of bile acids, SNA binding to B cells was reduced, indicating less α-2,6 or α-2,3-linked terminal galactose (Fig. 4B). In contrast, incubation of Raji-B cells with LCA and CDCA increased binding of AAL, ConA, and JAC (Fig. 4B). Together, these results demonstrate the differential expression of glycogenes and glycan structures following exposure to bile acids. Thus,

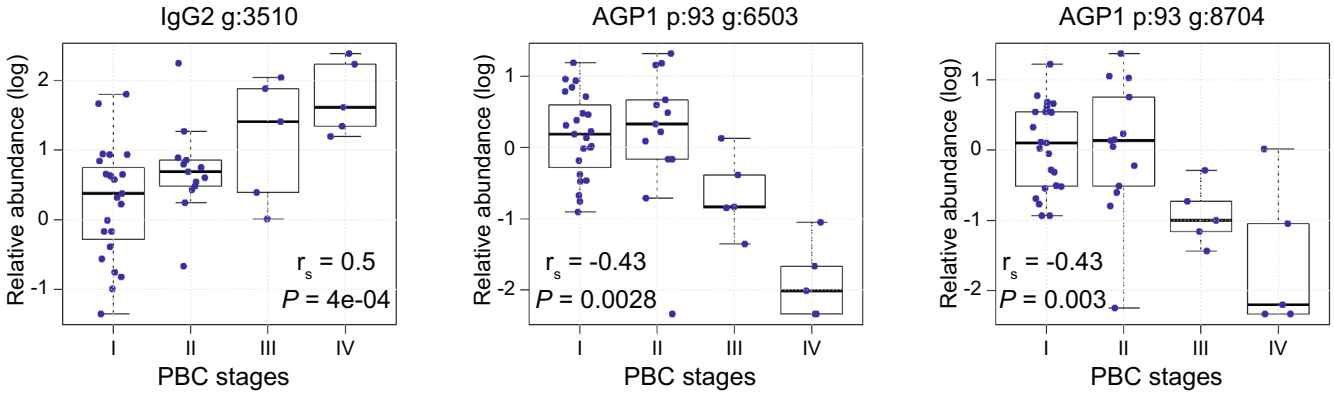
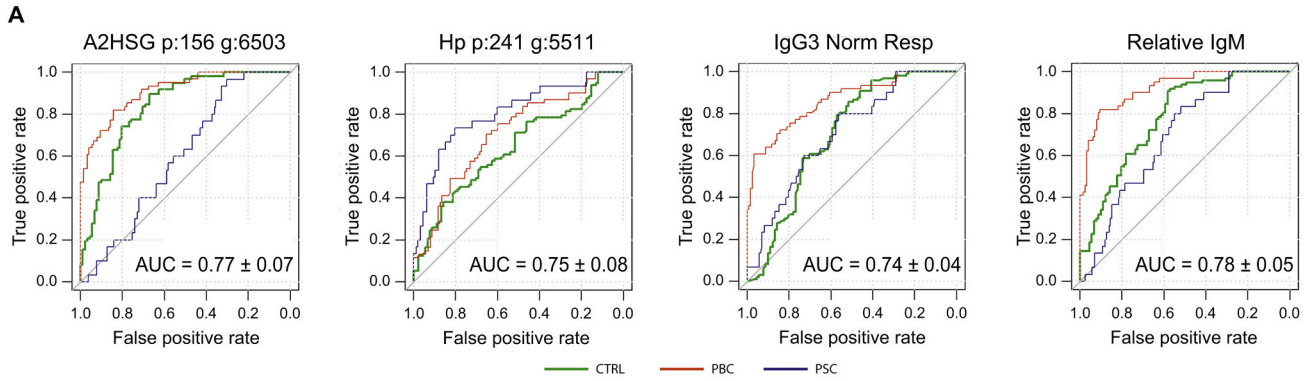
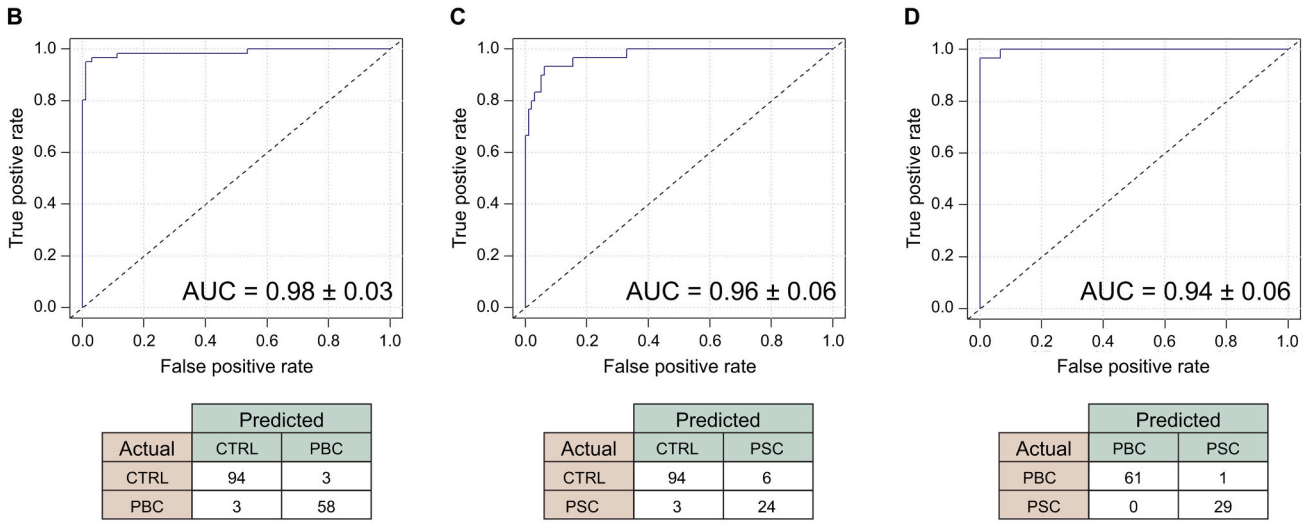


Fig. 2. Analytes altered by PBC stages. Log relative glycan abundance is graphed versus PBC stage and Spearman's correlation coefficients (r_s) and significance are calculated. For each stage the relative glycan abundance is represented as a box-and-whisker plot. The upper and lower bars connected to each box indicate the boundaries of the normal distribution and the upper and lower box edges mark the first and third quartile boundaries within each distribution. The bold line within the box indicates the median value of the distribution.

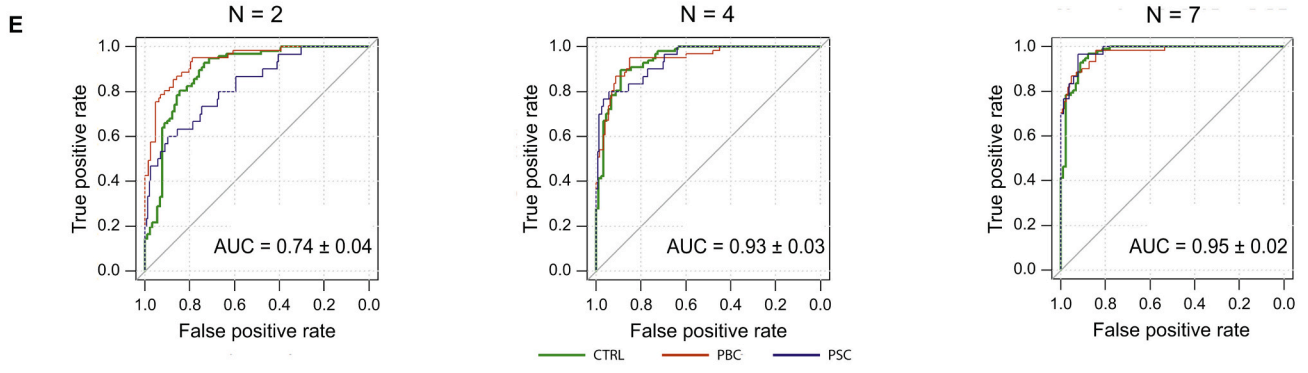




	Predicted	
Actual	CTRL	PBC
CTRL	94	3
PBC	3	58

	Predicted	
Actual	CTRL	PSC
CTRL	94	6
PSC	3	24

	Predicted	
Actual	PBC	PSC
PBC	61	1
PSC	0	29

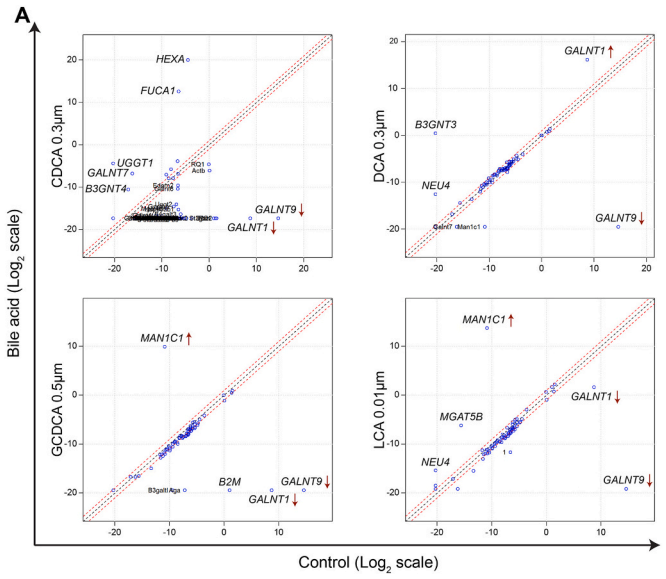


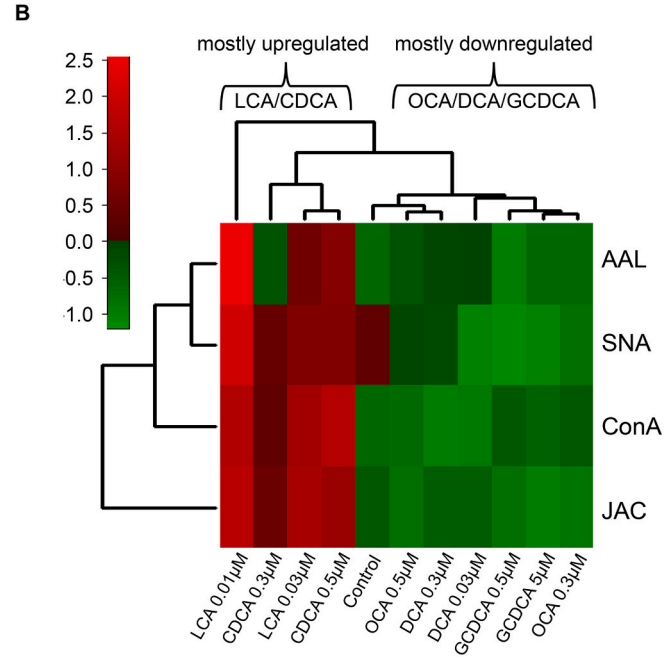
	Predicted		
Actual	CTRL	PBC	PSC
CTRL	86	9	13
PBC	9	50	3
PSC	2	2	14

	Predicted		
Actual	CTRL	PBC	PSC
CTRL	88	9	8
PBC	7	51	3
PSC	2	1	19

	Predicted		
Actual	CTRL	PBC	PSC
CTRL	90	6	3
PBC	3	53	3
PSC	4	2	24

Fig. 3. Classifiers for PBC vs PSC vs CTRL. (A) Receiver operating characteristic (ROC) curves of the 4 best single analyte classifiers for distinguishing between all diagnostic groups (PBC, PSC and CTRL). AUC includes 5-fold CV. A comprehensive list of the single analyte classifiers can be found in Table S2. (B) ROC curve for binomial multi-analyte classifier for distinguishing PBC from CTRL. Classifier is comprised of five glycoforms (ApoC3 p:74 g:1102, Hp p:207 g:11904, IgG2 g:5510, TF p:432 g:6502, and A1AT p:70 g:5402) and 1 plasma protein (relative IgM). (C) ROC curve for binomial multi-analyte classifier for distinguishing PSC from CTRL. Classifier is comprised of four glycoforms (Hp p:207 g:11904, A1AT p:70 g:5402, Hp p:241 g:5511, Hp p:184 g:6512). (D) ROC curve for binomial multi-analyte classifier for distinguishing PBC from PSC. Classifier is comprised of two glycoforms (AGP1 p:33 g:6501, AGP1/2 p:72MC g:7614), 1 plasma protein (relative IgM), and age as variables. (B to D) Confusion matrices representing classifier performance against the entire dataset are listed below their corresponding ROC curves. Five-fold cross validation results are displayed within the ROC curves. (E) Confusion matrices for multi-analyte classifiers designed to distinguish PBC, PSC, and controls from one another. The final model is represented by $n = 4$, made up of three glycoforms (Hp p:207 g:11904, A1AT p:70 g:5402, Hp p:241 g:5511) and 1 plasma protein (relative IgM), as increasing n further leads to only modest increases in classifier accuracy. A list of the remaining ROC curves with different number of predictors from $n = 1$ to $n = 13$ can be found in Fig. S4B.





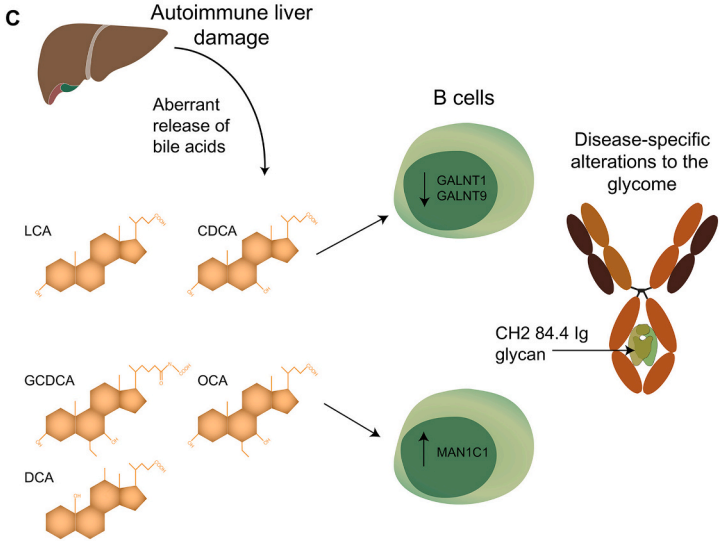


Fig. 4. Bile acids alter the glycosylations on B cells. (A) Raji B cells were incubated in triplicate with bile acids (CDCA, OCA, GCDCA, LCA, and DCA) at indicated concentrations chosen to represent normal and disease-relevant elevations in bile acids. The cells were harvested 48 h later and RNA extracted. A glycogene quantitative real-time PCR array was then used to characterize glycogene expression, which revealed differential expression of several genes including GALNT1 and GALNT9, which were consistently downregulated. The dotted lines present in each panel represent the 3-fold threshold. The remainder of the plots are shown in Fig. S5. (B) Raji B cells were treated in an identical fashion to part A. After 48 h cells were harvested and stained with a panel of fluorescently labeled lectins (AAL, ConA, JAC, SNA) and analyzed by flow cytometry. The strength of staining (MFI) of the bile acid cultured B cells and control cultured cells were used to calculate SMD values, which are presented graphically as a heatmap. (C) Schematic of glycan alterations in response to bile acids.

alterations in glycoforms in the setting of PBC and PSC may be influenced by both intra- and extracellular factors, the latter being the result of disease-associated environmental changes (Fig. 4C).

4. Discussion

Several studies have demonstrated glycan alterations in the setting of autoimmunity [4,28–39]. Some of these alterations occur at the conserved N-glycan glycosylation site within the Fc region of IgG, which is known to modulate Ig effector function [4]. All IgG subclasses (IgG1-4) have this same conserved immunoglobulin constant heavy chain domain 2 (CH2)-84.4 glycosylation site [29]. Sialylated N-glycans are anti-inflammatory at this site [4,40–44] and G0 glycans, which terminate in GlcNAc residues, are proinflammatory [4,28,38,45].

To characterize (CH2)-84.4 and other glycosylations in a site-specific fashion we have adopted MRM MS for glycan biomarker research and discovery. MRM MS allows for rapid characterization of site-specific glycosylations but requires predetermined knowledge of the glycopeptides electrospray ionization behavior, their collision induced dissociation patterns, as well as their retention times. These were previously established in our prior studies. For example, as a prelude to the current study we have mapped out the relative abundances of the 159 most common glycopeptides in the plasma of 97 healthy volunteers [46]. We have also extensively characterized the glycan alterations associated with age and gender [46], which allowed these factors to be accounted for in the current study.

In the current study, we characterized the site-specific glycan alterations associated with primary biliary cirrhosis, primary sclerosing cholangitis and healthy controls. According to the Altered Glycan Theory of Autoimmunity, each autoimmune disease will have a unique glycan signature [4]. Focusing on two related autoimmune diseases, PBC and PSC, allowed us to evaluate the utility of glycans as biomarkers of human disease. PBC was chosen as the prototypical disease for our initial autoimmune glycan biomarker study because it is not usually treated with immunosuppressive medications. Such medications may confound results if they have the ability to alter the glycosylation of plasma proteins [47]. PSC was chosen as a comparison group because it is also a liver-specific autoimmune disease but, unlike PBC, a PSC-specific blood test does not exist, making the discovery of a novel PSC serum biomarker of significant clinical value.

After adjusting for covariates and accounting for FDR, 61 of the monitored glycoforms were found to be significantly associated with PBC, 56 with PSC, and 47 were differentially expressed between PBC and PSC. Thus, while the levels of some glycans appeared to be disease-specific, others were more indicative of an autoimmune disease state (either PBC or PSC). For example, the 5411 IgG1 glycan was significantly decreased in both PBC and PSC patients, a finding we predicted *a priori* given that this glycan contains a terminal galactose residue and has both sialic acid and fucose decorations, which are thought to be anti-inflammatory [4]. Although there was some overlap, the overall glycosylation profiles associated with PBC and PSC were distinct from each other and from the profile that was associated with age. For example, the IgM Asn-209 glycosylation 5411 was significantly (FDR = 7.2e-06) altered by age but was not associated with either PBC or PSC (FDR > 0.05) [46]. Another important finding was that the relative abundance of some glycans was linked to disease stage. Thus, in the setting of autoimmunity there are many forces driving glycan alterations. At one end are the soluble cytokines and other inflammatory mediators pushing differential glyco-enzyme gene expression. At the other end are the influences resulting from the aftermath of autoimmune tissue destruction. These destructive forces can alter glycosylation by several means. For example, if the cells secreting the glycoproteins are damaged, this may be reflected in the glycovariants that they produce. Another possibility is that the autoimmune-mediated tissue destruction may release environmental elements that in turn induce changes in glycosylation. This possibility is supported by the ability of different bile acids (identical to those released by a damaged liver) to induce alterations in the glycosylation profile of Raji B cells. In addition to altering the plasma glycome, it is likely that elevations in bile acid levels will alter the glycosylation of epithelial cell integral membrane proteins, creating an altered-self, potentially recognizable by the immune system as foreign, thereby initiating an autoreactive immune response.

Apart from changes in glycosylation, it is becoming increasingly apparent that some autoimmune diseases are strongly associated with a particular Ig class or subclass. Prototypic examples include the IgG4-mediated diseases, pemphigus foliaceus and autoimmune pancreatitis [48,49]. Of relevance to our study, PBC has been previously reported to be associated with mitochondrial-specific IgM and IgG3 autoantibodies [50]. Our results demonstrate that relative IgM and IgG3 are significantly elevated in patients with PBC, which matches the anti-mitochondrial Igs' class/subclass and makes sense from an immunological perspective, as IgM and IgG3 are thought to work in concert during inflammatory immune responses [51]. Of note, IgM is negatively associated with age [46], which is the direct opposite of that observed in PBC, a disease mainly presenting in middle-aged to elderly individuals. The ability to monitor relative abundances is unique to our novel MRM mass spectrometry approach, which uses robustly quantified non-glycosylated common peptides to normalize glycoform abundances. This technique also normalizes Ig abundances across different isotypes (e.g. IgG1-4) and our results demonstrate the superior sensitivity of relative as compared to total Ig concentrations, which was an *a priori* prediction based upon the large variation in "normal" Ig concentrations [4].

In addition to being the first report of glycan alterations occurring in the setting of PBC and PSC, our study offers a number of advantages over prior analyses: 1) the glycan quantification was site-specific across multiple plasma proteins including different Ig classes and subclasses, not just a single protein or released glycans, as is the case in other publications; 2) the MRM approach eliminated the need for additional protein purification or chemical processing, which allowed for large patient cohorts to be rapidly characterized; 3) the analysis was precise, rapid, and automated for high throughput; 4) it required only 2 μl of serum or plasma and little sample preparation, while current techniques require several mL of blood to quantitate Ig levels; and 5) in addition to total protein quantification, the technique provided the relative abundance of each glycoform, making it more suitable for biomarker research and discovery. For these reasons, the development of this approach as a clinical diagnostic tool is very appealing, especially when compared to its more labor-intensive alternatives. However, the most unique aspect of our study is that it successfully established multiple-analyte classifiers capable of differentiating one autoimmune disease from another and patients with autoimmunity from healthy controls. To date, several studies have demonstrated glycan alterations in the setting of autoimmunity, but none have investigated the usefulness of site-specific glycosylations and multi-analyte classifiers as disease-specific biomarkers. The MRM technology that we employed in this current study is rapidly evolving and more site-specific glycosylations are being incorporated every month. This will dramatically increase the accuracy of our disease-specific classifiers. We anticipate that in the near future glycan analysis will become integral to the diagnosis and management of human diseases, especially diseases of the immune system and cancer.

Author contributions

E.M. designed the study, performed the statistical analyses, supervised the study, interpreted data and wrote the manuscript. A.A.M. designed the study, performed the statistical analysis, analyzed data and wrote the manuscript. D.P, M.J.K., G.X., L. R. R. performed the experiments and analyzed data. K.K. performed the statistical analysis and analyzed data and helped in writing of the manuscript. Q.H., Q.L., and Y. X. performed experiments and analyzed data. C.B., A.I.M. and P.L. supervised the study and wrote the manuscript. Y.J.Y.W., W.L, and N.N.L.

helped in writing of the manuscript and analyzed data. G.L. performed experiments and wrote the manuscript. C.B.L. designed the study, supervised the study and wrote the manuscript. All authors reviewed, edited and approved the manuscript prior to submission.

Data availability

All data are provided in the manuscript and the Supplementary Materials.

Funding

This work was supported by career awards from the Howard Hughes Medical Institute, Burroughs Wellcome Fund, Martin and Dorothy Charitable Spatz Foundation, and NIH Director's New Innovator Award (DP2OD008752) (to E.M.).

Declaration of Competing Interest

The authors declare no competing interests.

Appendix. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clim.2021.108825.

References

[1] R. Apweiler, H. Hermjakob, N. Sharon, On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, Biochim. Biophys. Acta 1473 (1999) 4–8.
[2] R.D. Cummings, The repertoire of glycan determinants in the human glycome, Mol. BioSyst. 5 (2009) 1087–1104.
[3] Transforming Glycoscience, A Roadmap for the Future, Washington (DC), 2012.
[4] E. Maverakis, K. Kim, M. Shimoda, M.E. Gershwine, F. Patel, R. Wilken, S. Raychaudhuri, L.R. Ruhaak, C.B. Lebrilla, Glycans in the immune system and the altered glycan theory of autoimmunity: a critical review, J. Autoimmun. 57 (2015) 1–13.
[5] Q. Hong, L.R. Ruhaak, C. Stroble, E. Parker, J. Huang, E. Maverakis, C.B. Lebrilla, A method for comprehensive glycosite-mapping and direct quantitation of serum glycoproteins, J. Proteome Res. 14 (2015) 5179–5192.
[6] A.C. Li, D. Alton, M.S. Bryant, W.Z. Shou, Simultaneously quantifying parent drugs and screening for metabolites in plasma pharmacokinetic samples using selected reaction monitoring information-dependent acquisition on a QTrap instrument, Rapid Commun. Mass Spectro.: RCM 19 (2005) 1943–1950.
[7] J.F. Xiao, B. Zhou, H.W. Ressom, Metabolite identification and quantitation in LC-MS/MS-based metabolomics, Trends Anal. Chem.: TRAC 32 (2012) 1–14.
[8] N.R. Kitteringham, R.E. Jenkins, C.S. Lane, V.L. Elliott, B.K. Park, Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 877 (2009) 1229–1239.
[9] S. Gallien, E. Duriez, B. Domon, Selected reaction monitoring applied to proteomics, J. Mass Spect.: JMS 46 (2011) 298–312.
[10] L.R. Ruhaak, C.B. Lebrilla, Applications of multiple reaction monitoring to clinical glycomics, Chromatographia 78 (2015) 335–342.
[11] K.D. Lindor, M.E. Gershwin, R. Poupon, M. Kaplan, N.V. Bergasa, E.J. Heathcote, D. American Association for Study of Liver, Primary biliary cirrhosis, Hepatology 50 (2009) 291–308.
[12] R. Chapman, J. Fevery, A. Kalloo, D.M. Nagorney, K.M. Boberg, B. Shneider, G. J. Gores, D. American Association for the Study of Liver, Diagnosis and management of primary sclerosing cholangitis, Hepatology 51 (2010) 660–678.
[13] Q. Hong, C.B. Lebrilla, S. Miyamoto, L.R. Ruhaak, Absolute quantitation of immunoglobulin G and its glycoforms using multiple reaction monitoring, Anal. Chem. 85 (2013) 8585–8593.
[14] S. Miyamoto, C.D. Stroble, S. Taylor, Q. Hong, C.B. Lebrilla, G.S. Leiserowitz, K. Kim, L.R. Ruhaak, Multiple reaction monitoring for the quantitation of serum protein glycosylation profiles: application to ovarian cancer, J. Proteome Res. 17 (2018) 222–233.
[15] V. R Foundation for Statistical Computing, Austria, R Development Core Team, R: A Language and Environment for Statistical Computing, 2008, p. 2008.
[16] M.P.J. van der Loo, Extremevalues, an R package for outlier detection in univariate data, in: R Package Version 2.1, 2014.
[17] J. Fox, S. Weisberg, An {R} Companion to Applied Regression, Second Ed, Sage, Thousand Oaks, CA, 2011.
[18] P. Royston, D.G. Altman, Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling, Appl. Stat. 43 (1994) 429–467.
[19] T. Lumley, A. Miller, Leaps: Regression Subset Selection. R Package Version 3.0, 2017.
[20] M. Kuhn, J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, T. Hunt, caret: Classification and Regression Training. R package version 6.0–76, 2017.
[21] G. Heinze, M. Ploner, logistf: Firth's Bias-Reduced Logistic Regression. R package version 1.22, 2016.
[22] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, Fourth ed., Springer, New York, 2002.
[23] A.D. Cullmann, HandTill2001: Multiple Class Area under ROC Curve. R Package Version 0.2–12, 2016.
[24] R. Brant, Assessing proportionality in the proportional odds model for ordinal logistic regression, Biometrics 46 (1990) 1171–1178.
[25] W. Viechtbauer, Conducting meta-analyses in R with the metafor package, J. Stat. Softw. 36 (2010) 1–48.
[26] J. Trottier, A. Białek, P. Caron, R.J. Straka, J. Heathcote, P. Milkiewicz, O. Barbier, Metabolomic profiling of 17 bile acids in serum from patients with primary biliary cirrhosis and primary sclerosing cholangitis: a pilot study, Dig. Liver Dis. 44 (2012) 303–310.
[27] L.N. Bell, J. Wulff, M. Comerford, R. Vuppalanchi, N. Chalasani, Serum metabolic signatures of primary biliary cirrhosis and primary sclerosing cholangitis, Liver Int. 35 (2015) 263–274.
[28] R.B. Parekh, R.A. Dwek, B.J. Sutton, D.L. Fernandes, A. Leung, D. Stanworth, T. W. Rademacher, T. Mizuochi, T. Taniguchi, K. Matsuta, et al., Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG, Nature 316 (1985) 452–457.
[29] R.B. Parekh, I.M. Roitt, D.A. Isenberg, R.A. Dwek, B.M. Ansell, T.W. Rademacher, Galactosylation of IgG associated oligosaccharides: reduction in patients with adult and juvenile onset rheumatoid arthritis and relation to disease activity, Lancet 1 (1988) 966–969.
[30] J.S. Moore, X. Wu, R. Kulhavy, M. Tomana, J. Novak, Z. Moldoveanu, R. Brown, P. A. Goepfert, J. Mestecky, Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals, Aids 19 (2005) 381–389.
[31] M. Holland, H. Yagi, N. Takahashi, K. Kato, C.O. Savage, D.M. Goodall, R. Jefferis, Differential glycosylation of polyclonal IgG, IgG-Fc and IgG-Fab isolated from the sera of patients with ANCA-associated systemic vasculitis, Biochim. Biophys. Acta 1760 (2006) 669–677.
[32] H. Homma, K. Tozawa, T. Yasui, Y. Itoh, Y. Hayashi, K. Kohri, Abnormal glycosylation of serum IgG in patients with IgA nephropathy, Clin. Exp. Nephrol. 10 (2006) 180–185.
[33] R. Saldova, L. Royle, C.M. Radcliffe, U.M. Abd Hamid, R. Evans, J.N. Arnold, R. E. Banks, R. Hutson, D.J. Harvey, R. Antrobus, S.M. Petrescu, R.A. Dwek, P. M. Rudd, Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG, Glycobiology 17 (2007) 1344–1356.
[34] M.H. Selman, E.H. Niks, M.J. Titulaer, J.J. Verschuuren, M. Wuhrer, A.M. Deelder, IgG fc N-glycosylation changes in Lambert-Eaton myasthenic syndrome and myasthenia gravis, J. Proteome Res. 10 (2011) 143–152.
[35] K. Kodar, J. Stadlmann, K. Klaamas, B. Sergeyev, O. Kurtenkov, Immunoglobulin G Fc N-glycan profiling in patients with gastric cancer by LC-ESI-MS: relation to tumor progression and survival, Glycoconj. J. 29 (2012) 57–66.
[36] M.H. Selman, S.E. de Jong, D. Soonawala, F.P. Kroon, A.A. Adegnika, A.M. Deelder, C.H. Hokke, M. Yazdanbakhsh, M. Wuhrer, Changes in antigen-specific IgG1 Fc N-glycosylation upon influenza and tetanus vaccination, Mol. Cellular Proteomics: MCP 11 (2012). M111 014563.
[37] L.R. Ruhaak, U.T. Nguyen, C. Stroble, S.L. Taylor, A. Taguchi, S.M. Hanash, C. B. Lebrilla, K. Kim, S. Miyamoto, Enrichment strategies in glycomics-based lung cancer biomarker development, Proteomics Clin. Appl. 7 (2013) 664–676.
[38] R. Parekh, D. Isenberg, G. Rook, I. Roitt, R. Dwek, T. Rademacher, A comparative analysis of disease-associated changes in the galactosylation of serum IgG, J. Autoimmun. 2 (1989) 101–114.
[39] A. Bond, A. Alavi, J.S. Axford, B.E. Bourke, F.E. Bruckner, M.A. Kerr, J.D. Maxwell, K.J. Tweed, M.J. Weldon, P. Youinou, F.C. Hay, A detailed lectin analysis of IgG glycosylation, demonstrating disease specific changes in terminal galactose and N-acetylglucosamine, J. Autoimmun. 10 (1997) 77–85.
[40] P. Bruhns, A. Samuelsson, J.W. Pollard, J.V. Ravetch, Colony-stimulating factor-1-dependent macrophages are responsible for IVIG protection in antibody-induced autoimmune disease, Immunity 18 (2003) 573–581.
[41] B. Tackenberg, I. Jelcic, A. Baerenwaldt, W.H. Oertel, N. Sommer, F. Nimmerjahn, J.D. Lunemann, Impaired inhibitory Fcgamma receptor IIB expression on B cells in chronic inflammatory demyelinating polyneuropathy, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 4788–4792.
[42] R.J. Hansen, J.P. Balthasar, Effects of intravenous immunoglobulin on platelet count and antiplatelet antibody disposition in a rat model of immune thrombocytopenia, Blood 100 (2002) 2087–2093.
[43] Y. Kaneko, F. Nimmerjahn, J.V. Ravetch, Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation, Science 313 (2006) 670–673.
[44] R.M. Anthony, F. Wermeling, M.C. Karlsson, J.V. Ravetch, Identification of a receptor required for the anti-inflammatory activity of IVIG, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 19571–19578.
[45] F. Listi, G. Candore, M.A. Modica, M. Russo, G. Di Lorenzo, M. Esposito-Pellitteri, G. Colonna-Romano, A. Aquino, M. Bulati, D. Lio, C. Franceschi, C. Caruso, A study of serum immunoglobulin levels in elderly persons that provides new insights into B cell immunosenescence, Ann. N. Y. Acad. Sci. 1089 (2006) 487–495.
[46] A.A. Merleev, D. Park, Y. Xie, M.J. Kailemia, G. Xu, L.R. Ruhaak, K. Kim, Q. Hong, Q. Li, F. Patel, Y.-J.Y. Wan, A.I. Marusina, I.E. Adamopoulos, N.N. Lal, A. Mitra, S. T. Le, M. Shimoda, G. Luxardi, C.B. Lebrilla, E. Maverakis, A site-specific map of

the human plasma glycome and its age and gender-associated alterations, Sci. Rep. 10 (2020) 17505.
[47] R. Saldova, J.E. Huffman, B. Adamczyk, A. Muzinic, J.J. Kattla, M. Pucic, M. Novokmet, J.L. Abrahams, C. Hayward, I. Rudan, S.H. Wild, A.F. Wright, O. Polasek, G. Lauc, H. Campbell, J.F. Wilson, P.M. Rudd, Association of medication with the human plasma N-glycome, J. Proteome Res. 11 (2012) 1821–1831.
[48] B. Rock, C.R. Martins, A.N. Theofilopoulos, R.S. Balderas, G.J. Anhalt, R.S. Labib, S. Futamura, E.A. Rivitti, L.A. Diaz, The pathogenic effect of IgG4 autoantibodies in endemic pemphigus foliaceus (fogo selvagem), N. Engl. J. Med. 320 (1989) 1463–1469.
[49] T. Kamisawa, N. Funata, Y. Hayashi, Y. Eishi, M. Koike, K. Tsuruta, A. Okamoto, N. Egawa, H. Nakajima, A new clinicopathological entity of IgG4-related autoimmune disease, J. Gastroenterol. 38 (2003) 982–984.
[50] L. Zhang, A.P. Weetman, D.R. Jayne, I. Turner, S.J. Yeaman, M.F. Bassendine, D. B. Oliveira, Anti-mitochondrial antibody IgG subclass distribution and affinity in primary biliary cirrhosis, Clin. Exp. Immunol. 88 (1992) 56–61.
[51] A.M. Collins, K.J. Jackson, A temporal model of human IgE and IgG antibody function, Front. Immunol. 4 (2013) 235.