

ARTICLE

DOI: 10.1038/s41467-017-00662-w

OPEN

A method to identify trace sulfated IgG N-glycans as biomarkers for rheumatoid arthritis

Jing-Rong Wang^{1,2}, Wei-Na Gao^{1,2}, Rudolf Grimm³, Shibo Jiang^{4,5}, Yong Liang^{1,6}, Hua Ye⁷, Zhan-Guo Li⁷, Lee-Fong Yau^{1,2}, Hao Huang^{1,2}, Ju Liu⁸, Min Jiang⁸, Qiong Meng^{1,2}, Tian-Tian Tong^{1,2}, Hai-Hui Huang⁶, Stephanie Lee⁹, Xing Zeng¹⁰, Liang Liu¹ & Zhi-Hong Jiang^{1,2}

N-linked glycans on immunoglobulin G (IgG) have been associated with pathogenesis of diseases and the therapeutic functions of antibody-based drugs; however, low-abundance species are difficult to detect. Here we show a glycomic approach to detect these species on human IgGs using a specialized microfluidic chip. We discover 20 sulfated and 4 acetylated N-glycans on IgGs. Using multiple reaction monitoring method, we precisely quantify these previously undetected low-abundance, trace and even ultra-trace N-glycans. From 277 patients with rheumatoid arthritis (RA) and 141 healthy individuals, we also identify N-glycan biomarkers for the classification of both rheumatoid factor (RF)-positive and negative RA patients, as well as anti-citrullinated protein antibodies (ACPA)-positive and negative RA patients. This approach may identify N-glycosylation-associated biomarkers for other autoimmune and infectious diseases and lead to the exploration of promising glycoforms for antibody therapeutics.

¹ State Key Laboratory of Quality Research in Chinese Medicine, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, China.

² Macau Institute for Applied Research in Medicine and Health, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, China.

³ Agilent Technologies, 5301 Stevens Creek Blvd, Santa Clara, CA 95051, USA. ⁴ Key Laboratory of Medical Molecular Virology of Ministries of Education and Health, Basic Medical College, Fudan University, Shanghai 200032, China. ⁵ Lindsley F. Kimball Research Institute, New York Blood Center, NY 10065, USA.

⁶ Faculty of Information Technology, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, China. ⁷ Department of Rheumatology and Immunology, Peking University People's Hospital, 11 Xizhimen South Street, Beijing 100044, China. ⁸ Division of Rheumatology, Jiujiang First People's Hospital, Taling North Road 48, Jiujiang 332000, China. ⁹ Agilent Technologies Hong Kong Ltd., Suite 2603, 26/F, AXA Tower, Landmark East, Kwun Tong, Hong Kong, China. ¹⁰ Guangdong Provincial Hospital of Chinese Medicine, Dade Road 111, Guangzhou 510120, China. Correspondence and requests for

materials should be addressed to L.L. (email: liu@must.edu.mo) or to Z.-H.J. (email: zhjiang@must.edu.mo)

Glycans can affect the clearance of pathogens and toxins as the specificity of IgG variable regions are coupled to Fc-mediated cellular functions which associate closely with Fc-linked glycans^{1, 2}. In particular, IgG glycoforms have been used as rheumatoid arthritis (RA), and the prediction of immune responses^{2–8}. Glycosylation also has an important effect on the potency and safety of therapeutic antibodies, owing to effects on activity and possibly also stability, solubility, clearance and immunogenicity of antibody-based drugs, such as intravenous immunoglobulin (IVIG) and monoclonal antibodies (mAb)⁹. Therefore, the identification of functional N-glycans is a promising area of biomedical research^{1, 10}.

Glycans are encoded in a complex dynamic network of hundreds of genes that participate in the biosynthetic pathway of protein glycosylation, resulting in the generation of extremely complicated and variable glycosylation profiles, even for single glycoproteins^{11, 12}. Such heterogeneity of glycan structures means that individual glycoforms may have very low abundance. According to Go et al.¹³, complete glycosylation analysis is a great challenge due to the complexity of samples, wide dynamic range of glycosylation level and structure heterogeneity of glycans. Notably, many low-abundance and trace N-glycans are biologically important, whereas high-abundance N-glycans generally function as ‘safe switches’^{11, 14, 15}. Acidic N-glycans with anionic residues, such as sialic acid, sulfate, and phosphate groups, are examples of such low-abundance but biologically important IgG species^{16–18}. For example, IgGs with sialylated N-glycans account for a small portion of pooled IgGs, but have anti-inflammatory activity via binding to dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin (DC-SIGNR), rather than Fcγ receptors^{9, 16, 19}. In addition, the ionization efficiency of acidic N-glycans is generally lower than that of neutral N-glycans, resulting in much lower signal intensities^{18, 20, 21}. Therefore, a comprehensive glycomic approach to detect low-abundance and difficult-to-detect, but biologically important, species is required.

Although mass spectrometers with enhanced sensitivity exist, the overall increased signal intensity can lead to elevated ion-suppression/interference arising from high-abundance neutral N-glycans/matrix¹¹. Chromatographic enrichment is essential for improving the detection sensitivity of low-abundance acidic N-glycans and glycoproteins; however, it has not been done on a micro scale¹¹. To microminaturize and automate the process, we design a microfluidic chip in which a titanium dioxide (TiO₂) column enriches acidic N-glycans and a porous graphitized carbon (PGC) column chromatographically separates N-glycans. This design enables the comprehensive profiling and accurate quantification of N-glycans, including extremely low-abundance N-glycans, and particularly the acidic species.

Results

Design and workflow of TiO₂-PGC chip. The multilayered polymeric microfluidic chip with TiO₂-based enrichment architecture was first described in 2008 as a two dimensional HPLC separation device for phosphopeptide analysis²². This design enabled the integration of two independent separation functions on a single microfluidic device, where each separation function was on a different layer of the device. To develop a feasible and integrated glycomic approach, we have drawn on the multilayer chip concept and expanded it to the design and fabrication of a novel microfluidic chip, TiO₂-PGC chip. TiO₂ has a high affinity for negatively charged molecules^{23–26}. Accordingly, the rapid and highly selective on-chip enrichment of acidic N-glycans can be achieved using the high selectivity of TiO₂ toward negatively charged N-glycans. Thus the TiO₂-PGC chip is particularly designed for acidic N-glycans enrichment and analysis. It contains a ‘sandwich-like’ enrichment column composed of TiO₂ and PGC (Fig. 1a). This enrichment column is connected to analysis column with PGC stationary phase, ending in an integrated electro-spray tip. The chip cube allows mounting and positioning of the chip tip.

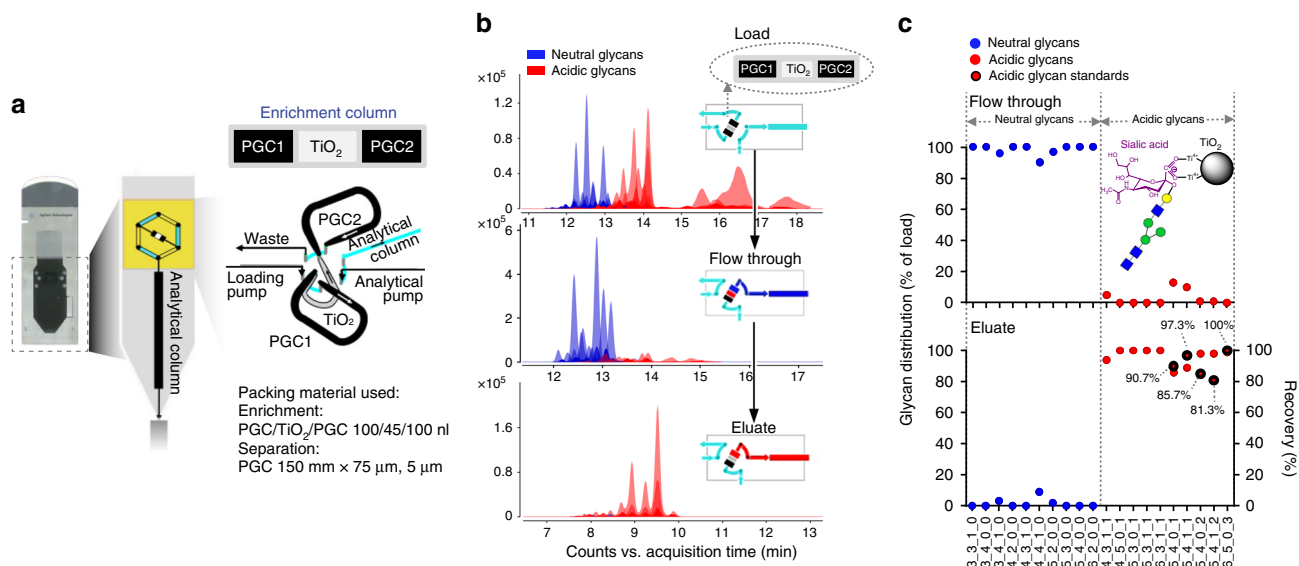


Fig. 1 On-chip enrichment of acidic N-glycans. **a** Diagram of the specialized TiO₂-PGC chip. **b** Extracted compound chromatograms (ECC) of N-glycans detected in the load, flow-through (neutral N-glycans) and eluate fractions (acidic N-glycans) of the TiO₂ enrichment column as analyzed by liquid chromatography coupled with nano-electrospray ionization quadrupole time-of-flight mass spectrometry in the positive mode. **c** TiO₂ enrichment column selectively captures acidic N-glycans. Neutral N-glycans (blue dot) do not efficiently bind to TiO₂ and are recovered in the flow-through fraction, whereas acidic N-glycans are retained on TiO₂ during the load and flow-through steps and are eluted with the elution buffer. The enrichment recovery rates (numbers beside the dots) of five acidic N-glycan standards (red dots with black circles) are also given

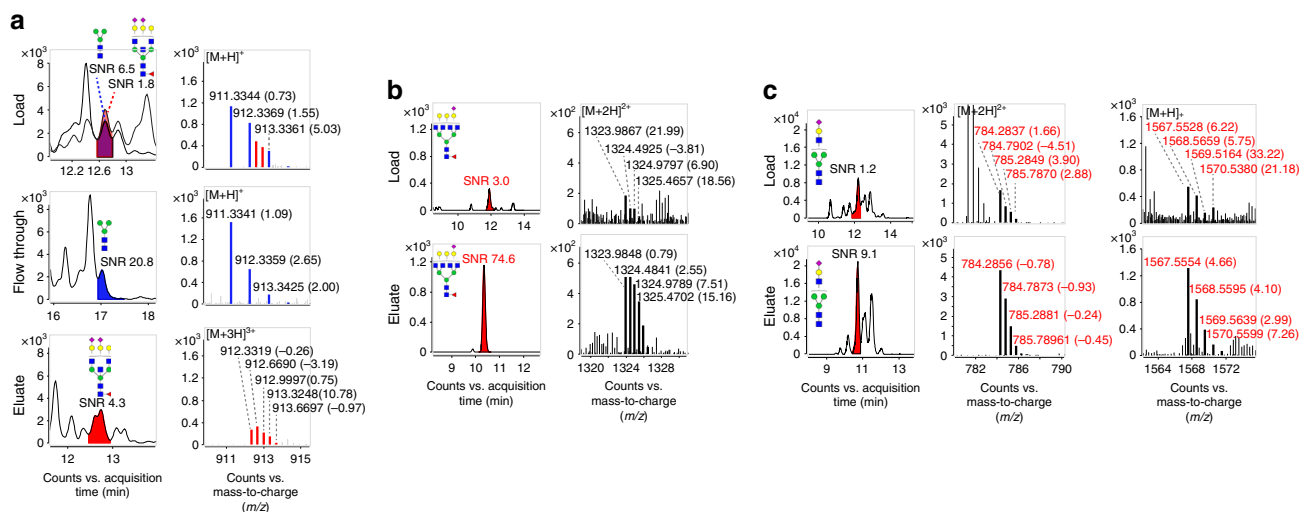


Fig. 2 The greatly improved detection of acidic N-glycans after enrichment on the TiO_2 -PGC chip. In **a**, the $[M+3H]^{3+}$ ions of an acidic N-glycan were overlapped with the isotopic ions of co-eluted neutral N-glycans and could not be assigned accurately. Using the TiO_2 -PGC chip, the acidic N-glycans were enriched efficiently and detected with an enhanced S/N and improved mass accuracy because of the removal of interference from neutral N-glycans. In **b**, the signal intensity as well as the S/N of acidic N-glycans were significantly enhanced after on-chip enrichment because of the reduced ion-suppression derived from neutral N-glycans/matrix. In **c**, with the removal of noise derived from neutral N-glycans/matrix, the acidic N-glycan mass accuracy was significantly enhanced, thus providing reliable evidence for the identification of acidic N-glycans. The number in the bracket beside each mass value indicates the mass error (p.p.m.)

During analysis, a micro valve is connected directly with the chip surface to create a zero dead volume via high-pressure seal. First, the sample is loaded onto the chip and all N-glycans are retained on the first PGC section. Second, using a linear gradient $\text{H}_2\text{O}/\text{acetonitrile}$ (ACN) as mobile phase, all N-glycans are eluted off the first PGC enrichment column. Neutral N-glycans pass the TiO_2 section and the second PGC enrichment column and get transferred onto the analytical column on which they are separated. During this process, acidic N-glycans are retained and enriched on the TiO_2 column. Third, by a plug of elution buffer (NH_3), those retained acidic N-glycans are desorbed and retained onto the second PGC enrichment column, and are transferred to the analytical column by subsequent gradient elution.

Improved detection of acidic N-glycans by on-chip enrichment.

The “sandwich” design of the enrichment column on the TiO_2 -PGC chip enables the enrichment of N-glycans on TiO_2 after a continuous “pre-fractionation” on PGC (PGC1, Fig. 1a), thus allowing for the enrichment of trace species from a complex N-glycan pool. Using a complicated mixture obtained from serum IgGs (Supplementary Fig. 1), we separated low-abundance acidic N-glycans from much more abundant neutral N-glycans with high selectivity (>80%, Fig. 1b, c) on this specialized chip. The coupling of this capability with the relatively high binding capacity of the enrichment column (Supplementary Fig. 2a) facilitated a broad and dynamic range of on-chip enrichment, as indicated by the quantitative enrichment of acidic species from total N-glycans of 0.03–0.5 μg IgG (Supplementary Fig. 2b). These features make this chip feasible for analyzing various biological samples in which the level of N-glycans may vary greatly.

Compared with the offline Strong Anion Exchange (SAX) method, the number of acidic N-glycans enriched using this online chip increased by about 50% (Supplementary Fig. 3a), and an overall improvement in reproducibility was demonstrated (Supplementary Fig. 3b–d). Notably, the relative abundances of N-glycans with varied numbers of sialic acids were conserved after enrichment, indicating that an enrichment bias was not

introduced and a “natural and genuine” profile of acidic N-glycans was achieved (Supplementary Fig. 4).

Using the TiO_2 -PGC chip, the detection of low-abundance acidic N-glycans was achieved with improved sensitivity because of the decreased ionization suppression derived from otherwise co-eluted neutral N-glycans and/or other matrix components, as demonstrated by the up to 25-fold increase in the signal-to-noise ratio (S/N) of acidic N-glycans after enrichment (Fig. 2, chromatograms). In addition, the online enrichment enabled enhanced mass accuracy and isotope distribution of the intact N-glycan’s ion because of the removal of neutral N-glycans and/or other molecules with similar molecular weights (Fig. 2, MS spectrum). Of note, the enrichment procedure facilitated reliable tandem Mass Spectrometry (MS/MS) measurements by removing isomeric and/or isobaric ions that can pass through the quadrupole to undergo fragmentation and compromise the fragmentation information.

Development of dual-mobile phase approach. In a subsequent step of method optimization, we attempted to use specific mobile phase to separately improve the detection of neutral and acidic N-glycans. To this end, we developed a “dual mobile phase” approach in line with the special work mode of TiO_2 -PGC chip. We examined different mobile phases for the analysis of neutral and acidic N-glycans individually, and then used optimal mobile phase for respective class of N-glycans. In the finalized dual-mobile phase approach, 1% formic acid (FA) in water (A1) and acetonitrile (ACN) (B1) were used as mobile phase of the first run (for the analysis of neutral and high mannose N-glycans), while 0.5% FA in water (adjusted to pH 3 by ammonia solution) (A2) and 1% FA in ACN (B2) were employed as the mobile phase of the second run (for the analysis of acidic N-glycans). The dual-mobile phase approach improved the detection sensitivity of three types of N-glycans by approximately 5-fold as compared to routinely used single mobile phase (Supplementary Fig. 5). This benefit, together with the improvements achieved by online enrichment, provided a firm basis for the detection of low-abundance acidic N-glycans, which cannot be achieved by

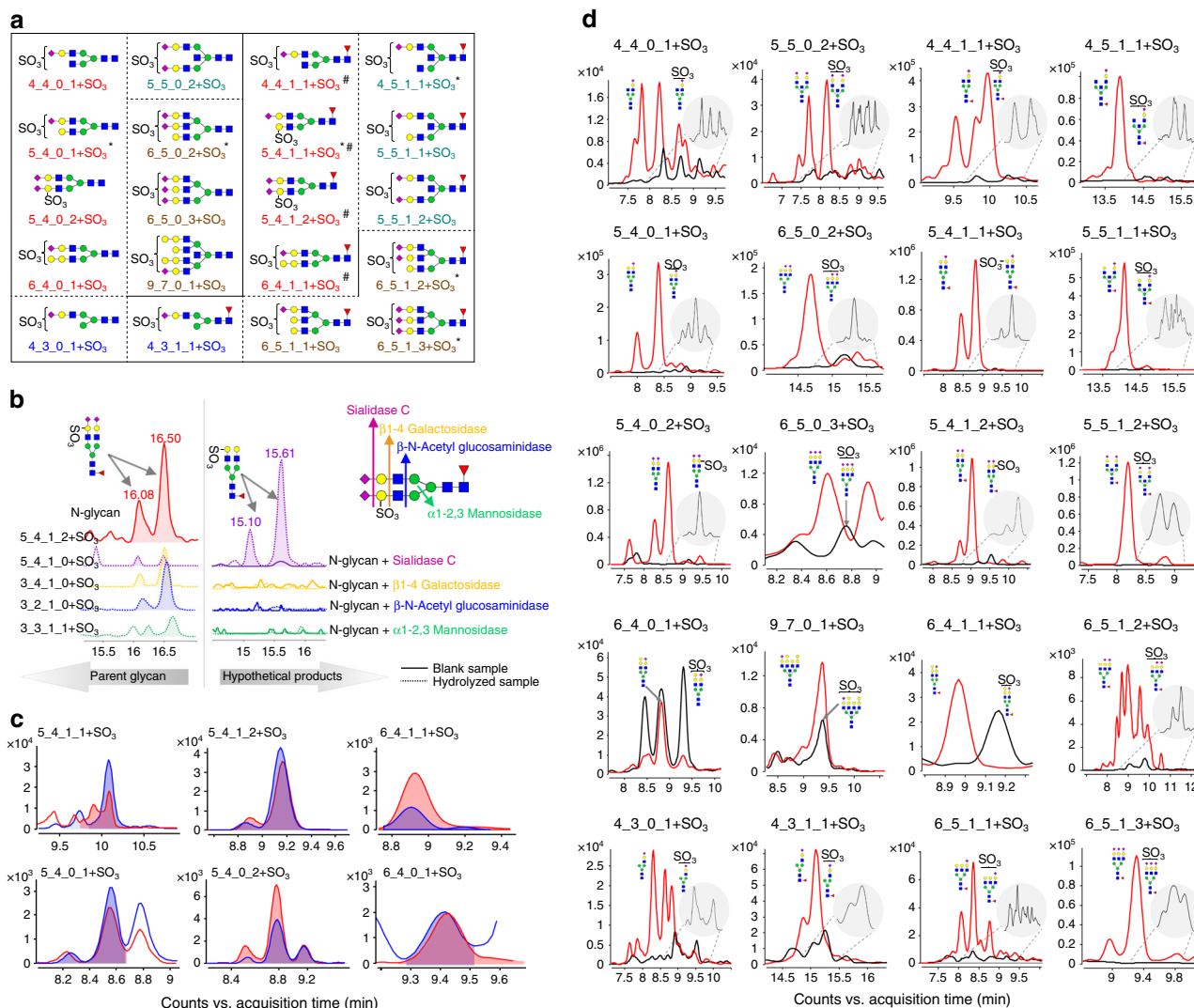


Fig. 4 Structures of sulfated N-glycans identified on human serum IgGs. **a** Structures of identified sulfated N-glycans in human serum IgG. (*represents sulfated N-glycans that have been reported from a human source; #represents sulfated N-glycans that have been reported on porcine thyroglobulin; and another 11 sulfated N-glycans have not been previously reported). **b** Four exoglycosidases, Sialidase C, β 1-4 Galactosidase, β -N-acetyl glucosaminidase, and α 1-2, 3 mannosidase, were employed to hydrolyze the sulfated N-glycans to determine site of the sulfate group. **c** The retention times and peak patterns of sulfated N-glycans in human IgGs (red) were compared with those in porcine thyroglobulin (blue), which has been well characterized. **d** The retention time of each sulfated N-glycan (black) is 0.5 min later than its corresponding parent N-glycans (red), and the signal intensity is 3-fold to 200-fold lower than that of the non-sulfated N-glycans

simply enhancing the overall signal intensity. As such, the dual-mobile phase approach was applied to both HPLC chip-QTOF MS system for N-glycan profiling and HPLC chip-QQQ-MS system for quantitative N-glycan analysis.

Comprehensive N-glycan profile of human serum IgGs. The specialized TiO₂-PGC chip, combined with dual-mobile phase method, facilitated an integrated glycomic approach that enabled the most comprehensive profiling of N-glycans on IgGs reported thus far. Using ~2.5 μ g serum polyclonal IgGs, a total of 444 N-glycans generated from 185 distinct compositions were identified in the current study, including 54 and 131 compositions of neutral and acidic N-glycans, representing 56% coverage of the recently developed theoretical N-glycan library for serum (331 compositions)²⁷. Notably, our identifications almost doubled the number of known neutral N-glycans of IgG (36 compositions) and almost quadrupled the number of acidic N-glycans previously identified on IgG (36 compositions; Fig. 3a)²⁸. The

newly identified N-glycans not only demonstrated a sharp increase in the detection capability of acidic N-glycans via the on-chip method but also suggested a remarkable and previously unknown structural diversity of acidic N-glycans on IgG.

The comprehensive profiling of acidic N-glycans resulted in an enlarged bioinformatics framework of N-glycans that can be attributed to the variation in saccharide compositions as indicated by multiple fucosylated (2–3) ternary, quaternary or pentanary N-glycan structures, and the modification of the N-glycan structures as represented by O-acetylation and sulfation which has never been observed on human serum glycoproteins (Fig. 3b, Supplementary Data 1, 2). More importantly, many sulfated N-glycans (36 structures arising from 20 distinct compositions) were discovered on IgGs for the first time, including 9 compositions that were previously found in recombinant human tissue plasminogen activator²⁹, human urine³⁰ and on porcine thyroglobulin^{31, 32}, respectively, and 11 compositions that had not been previously reported (Fig. 4a).

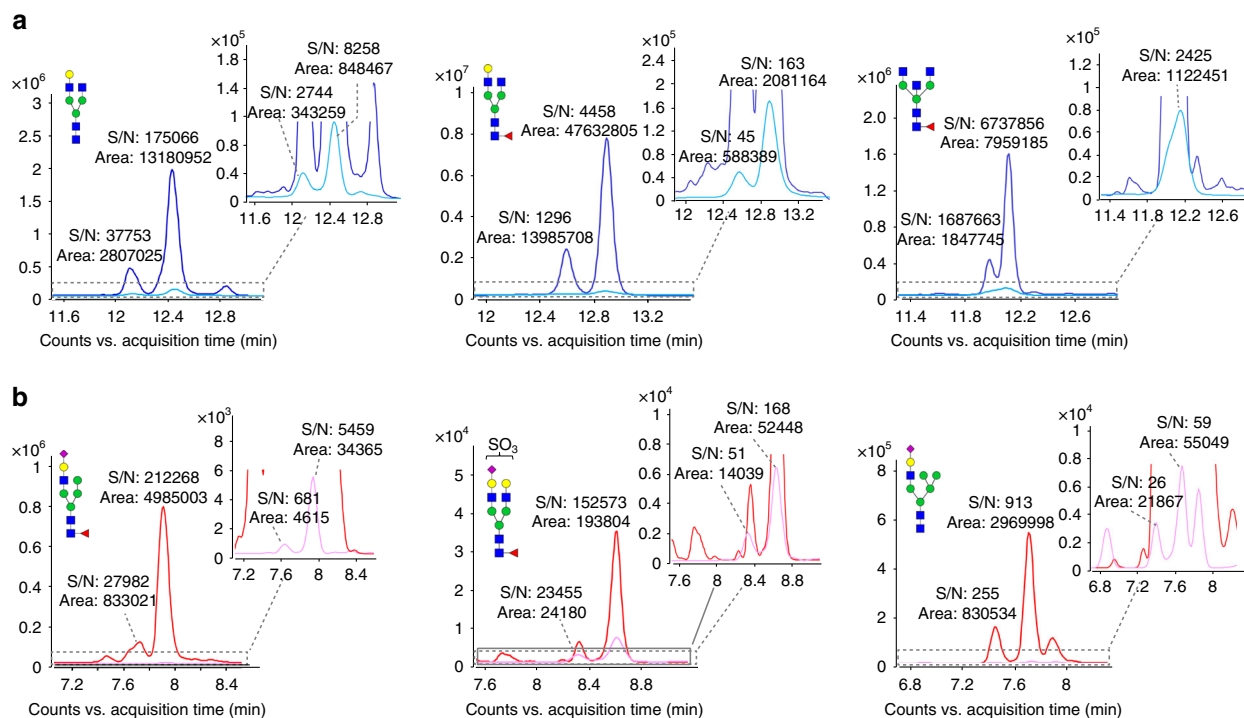


Fig. 5 Improved detection of N-glycans using the MRM method. N-glycans were detected using Q-TOF MS (pale blue or pale red) and QQQ MS (MRM mode, blue or red). The improved detection of representative neutral N-glycans Hex₄HexNAc₄, Hex₄HexNAc₄dHex₁ and Hex₃HexNAc₅dHex₁ are shown in **a**, and the improved detection of representative sialylated N-glycans Hex₅HexNAc₃dHex₁NeuAc₁, Hex₅HexNAc₄dHex₁NeuAc₁ + SO₃ and Hex₆HexNAc₃NeuAc₁ are shown in **b**. The signal intensities of N-glycans obtained using the MRM method were increased by up to 180-fold compared with those obtained with Q-TOF MS. Notably, the S/N of the acidic N-glycans obtained using MRM increased by ~10-fold to 1000-fold compared with that obtained with Q-TOF MS

The 20 sulfated N-glycan compositions were primarily assigned based on their high-resolution MS data (Supplementary Fig. 6, Supplementary Data 1). The presence of sulfate group was confirmed by the observation of diagnostic fragment ion originating from in-source neutral loss $[M+2H-SO_3]^{2+}$ or $[M+3H-SO_3]^{3+}$ (Supplementary Fig. 6c). The structures of the “net” N-glycan part were verified by high-resolution MS/MS data (Supplementary Data 1) for 19 compositions.

Of note, fragment ions with sulfate group were observed for 8 out of the 20 compositions in high-resolution MS/MS spectra, further demonstrating sulfation of these N-glycans. In addition, these fragments provided key information for the location of sulfate group. Take Hex₅HexNAc₄dHex₁NeuAc₂ + SO₃ (5_4_1_2 + SO₃) as an example, fragment ions corresponding to “[B7/Y4 + SO₃]⁺” (m/z 1485.4919), “[Y6 + SO₃]²⁺” (m/z 998.8133), “[B4 + SO₃]⁺” (m/z 899.2441), “[B3 + SO₃]⁺” (m/z 737.1942) and “[B4/Y7 + SO₃]⁺” (m/z 608.1388) were clearly observed in the MS/MS spectrum of the $[M+2H]^{2+}$ at m/z 1225.3985, suggesting that the sulfate group was located on galactose or N-acetyl glucosamine (Fig. 3c–f). This assignment was supported by exoglycosidase experiment which suggested the attachment of the sulfate group on galactose of α 1–3 branch (Fig. 4b).

In addition, a total of six sulfated N-glycans have been reported in porcine thyroglobulin (pTG), with NMR confirmed structures. We employed N-glycan mixture released from pTG as “authentic” sample and verified the identification of these sulfated N-glycans on IgGs (Fig. 4c). In addition, all of the sulfated N-glycans exhibited consistent retention time differences (typically +0.5 min) relative to their “parent” species, and their signal intensities were 3-fold to 200-fold lower than those of their

non-sulfated counterparts (Fig. 4d). The specific retention behavior can be employed as a supporting evidence for the identification of sulfated N-glycans. Sulfation constitutes a novel structural variation and represents another type of N-glycan microheterogeneity on IgG.

Quantitative glycomic approach by MRM. The multiple reaction monitoring (MRM) method was then used to quantify the N-glycans of IgGs (Supplementary Data 3), further improving the signal intensity of those low-abundance acidic N-glycans by ~1000-fold compared with that of TOF-MS (Fig. 5, Supplementary Table 1). In particular, the use of TiO₂ enrichment enabled the highly sensitive detection of acidic N-glycans mixed with complex neutral N-glycans (Supplementary Fig. 7). This quantitative method demonstrated much wider linearity (typically 250-fold to 2000-fold) compared with TOF-MS (generally 30-fold to 70-fold) (Supplementary Table 2). With the MRM method, acidic N-glycan signal intensities can be pointedly enhanced by increasing the dwelling time, thus significantly reducing ionization bias (Supplementary Fig. 8). This MS/MS-based method was further validated for its recovery rate and selectivity (Supplementary Table 3).

The glycome of human serum IgGs were quantitatively profiled by using the established MRM method. The result showed that both neutral and acidic N-glycans presented in IgG of human serum were predominantly in complex type, only less than 0.1% N-glycans were in high mannose type and about 1.8% neutral N-glycans and 0.2% acidic N-glycans were in hybrid type. Among the complex N-glycans, bi-antennary N-glycans occupied the highest portion (90% of overall N-glycans), and most of them were fucosylated (more than 75%) and sialylated (more than

40%), trace acidic N-glycans were modified by a sulfate group (about 1%; Fig. 6, Supplementary Data 4). The quantitative glycomic profiling revealed remarkable depth in the concentration of individual N-glycans on human serum IgGs.

N-glycan profile of serum IgGs from RA patients. To explore the potential biological role of acidic N-glycans, we extended our analysis to serum samples from RA patients. The entire N-glycomes of the serum IgGs of 90 RA patients and 57

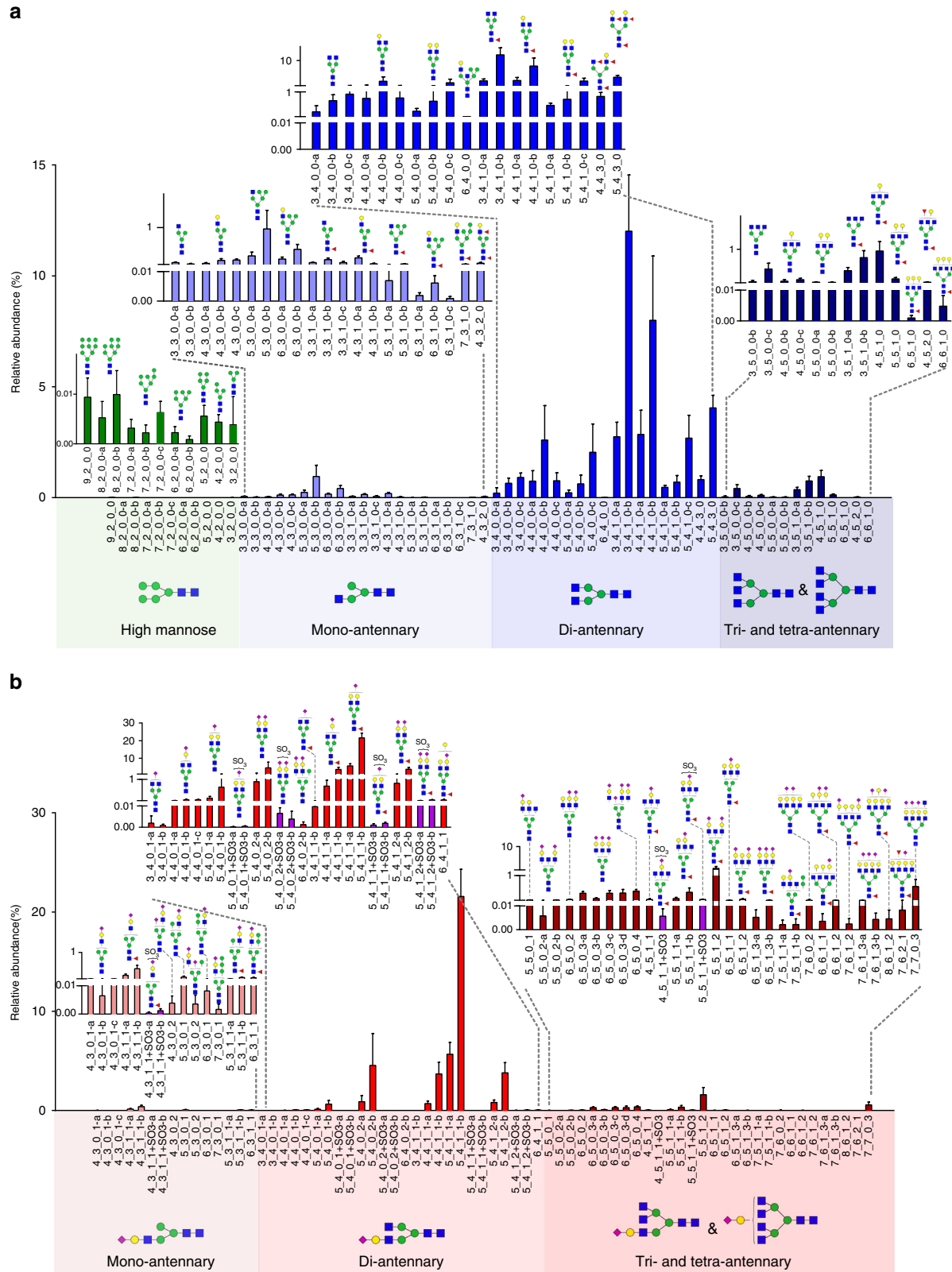


Fig. 6 Glycomic analysis of human serum IgGs using MRM method. It revealed a highly dynamic range of up to five orders of magnitude in the relative abundances of both neutral **a** and acidic **b** N-glycans, data are shown as mean \pm s.d

Table 1 Capacities of 21 N-glycan biomarkers for the classification of RA patients and healthy individuals

N-Glycan biomarkers	Training set (RA patient (n = 90) and healthy individuals (n = 57))			Validation set (RA patient (n = 187) and healthy individuals (n = 84))		
	AUC	Sensitivity (%)	Specificity (%)	AUC	Sensitivity (%)	Specificity (%)
<i>Neutral N-glycans</i>						
5_3_1_0-b	0.853	75	82	0.793	93	69
6_3_1_0-b	0.814	77	76	0.768	80	75
6_3_1_0-a	0.758	93	54	0.701	94	53
3_5_0_0-c	0.746	72	69	0.701	75	66
5_5_0_0-b	0.737	58	82	0.734	62	82
5_3_1_0-c	0.736	74	62	0.708	81	64
3_3_0_0-a	0.717	89	57	0.723	90	59
4_3_1_0-b	0.696	60	74	0.723	66	77
4_2_0_0	0.747	81	60	0.738	87	64
<i>Acidic N-glycans</i>						
8_6_1_2	0.803	82	71	0.779	84	73
4_3_1_1-a	0.791	81	62	0.786	60	92
4_3_0_1-c	0.758	86	56	0.705	87	59
5_5_0_2-b	0.740	70	69	0.705	73	69
6_5_0_2	0.724	84	53	0.705	75	68
7_5_1_1-b	0.690	79	60	0.705	82	62
5_4_0_1-a	0.657	93	32	0.539	93	25
4_4_1_1-a	0.601	68	56	0.661	53	76
<i>Sulfated N-glycans</i>						
5_4_1_1+SO ₃ -b (SGm2)	0.858	77	83	0.819	79	84
4_3_1_1+SO ₃ -b (SGm1)	0.814	75	80	0.793	80	79
5_4_1_2+SO ₃ -b	0.721	67	71	0.697	69	70
5_4_0_1+SO ₃ -a	0.683	81	52	0.664	85	53
All	0.869	82	78	0.804	82	79
SGm1 and SGm2	0.879	84	86	0.849	85	85

healthy individuals were primarily analyzed, and the relative abundances of individual N-glycans were employed as variations to predict the grouping of individuals. As shown in Supplementary Fig. 9, a general trend of “the more minor, the more widely changed” was observed for both neutral and acidic N-glycans.

For the identification of potential N-glycan biomarkers, we selected informative N-glycan biomarkers (attributes) by using a popular machine learning tool, Weka. The process was performed by employing attribute evaluator by which attribute subsets are assessed and search method with which the space of possible subsets are searched. We used correlation-based feature subset selection method (CfsSubsetEval) to select value subsets that correlate highly with the class value and lowly with each other. And we used the best-first search strategy to navigate attribute subsets in search space. We then fed the classification ready pre-processed data from the entire N-glycome of the serum IgGs of 90 RA patients and 57 healthy samples (training data) into Weka. As the results, 21 N-glycan biomarkers were selected. Subsequently, these 21 biomarkers were used to construct a diagnostic model by using Support Vector Machine (SVM) method individually.

The prediction models were measured for sensitivity and specificity as described in the method. For the individual 21 N-glycan biomarkers, the sensitivity values ranged from 58 to 93%, the specificity values varied from 32 to 83%, while the area under the curve (AUC) values ranged from 0.601 to 0.858 (Table 1). Receiver operating characteristic (ROC) curves were further constructed for the identified N-glycan biomarkers, which generated an overall sensitivity of 82 % and a specificity of 78% combined with an AUC of 0.869. Notably, four sulfated N-glycans were identified as biomarkers that distinguish RA patients from the controls (Supplementary Fig. 9c, f).

Among them, two sulfated N-glycans, 4_3_1_1+SO₃-b and 5_4_1_1+SO₃-b, yielded high individual AUCs of 0.814 and 0.858 respectively, as well as relatively high individual sensitivity (75% and 77%, respectively) and specificity (80% and 83% respectively) (Table 1 and Fig. 7a, b), indicating their particular potential value in the classification of RA. Thus, these two sulfated glycans were assigned as high-potential biomarkers for RA and termed as SGm1 and SGm2.

To maximize the ability of the N-glycan biomarkers for distinguishing RA patients from healthy individuals, we examined the classification abilities of pair biomarker combinations. Among all plausible combinations (Supplementary Data 5), combination of the two sulfated N-glycan biomarkers (SGm1 and SGm2) generated an optimal capacity for the classification of RA in training set (with AUC of 0.879, sensitivity of 84% and specificity of 86%).

We further examined the accuracy of individual biomarkers for the classification of RF-negative patients in training set (n = 19). All N-glycan biomarkers except for 4_2_0_0 were shown to have relatively high prediction capacity (accuracy ≥ 84%) with an overall accuracy of 89%. As ACPA is also ubiquitous in clinical care, we also examined the performance of the biomarkers for the classification of ACPA-negative patients (n = 15), as well as RF&ACPA-negative patients (n = 8). As shown in Table 2 and Fig. 7, the identified 21 biomarkers showed generally high prediction capacity for ACPA-negative and RF&ACPA-negative patients, with an overall prediction accuracy of 87 and 100% in training set respectively. The combination of SGm1 and SGm2 yielded prediction accuracy of 95, 93 and 95% for RF-negative, ACPA-negative and RF&ACPA-negative patients.

The sensitivity and specificity for the biomarkers are measured by identifying the cutoff point of the predicted probability that yielded the highest sum of sensitivity and specificity. For

Table 2 Prediction accuracy (%)^d of 21 N-glycan biomarkers for RF-negative and/or ACPA-negative patients

N-Glycan biomarkers	Training set			Validation set		
	R(-) ^a patients (n = 19)	A(-) ^b patients (n = 15)	R&A(-) ^c patients (n = 8)	R(-) ^a patients (n = 97)	A(-) ^b patients (n = 55)	R&A(-) ^c patients (n = 45)
<i>Neutral N-glycans</i>						
5_3_1_0-b	95	93	100	88	91	96
6_3_1_0-b	100	100	100	77	82	80
6_3_1_0-a	95	93	100	72	69	76
3_5_0_0-c	100	100	100	79	82	82
5_5_0_0-b	89	93	100	66	60	60
5_3_1_0-c	95	93	100	77	87	91
3_3_0_0-a	95	93	100	71	58	64
4_3_1_0-b	95	93	100	69	58	64
4_2_0_0	63	53	50	82	95	93
<i>Acidic N-glycans</i>						
8_6_1_2	89	87	75	80	75	78
4_3_1_1-a	95	93	88	75	75	76
4_3_0_1-c	89	93	88	78	82	84
5_5_0_2-b	100	100	100	71	65	76
6_5_0_2	89	93	88	72	73	76
7_5_1_1-b	89	87	75	77	78	82
5_4_0_1-a	100	100	100	59	51	60
4_4_1_1-a	89	87	88	59	51	60
<i>Sulfated N-glycans</i>						
5_4_1_1 + SO ₃ -b (SGm2)	95	93	88	84	82	82
4_3_1_1 + SO ₃ -b (SGm1)	100	100	100	76	78	73
5_4_1_2 + SO ₃ -b	95	93	100	64	51	60
5_4_0_1 + SO ₃ -a	84	80	88	70	82	82
All	89	87	100	81	80	80
SGm1 and SGm2	95	93	95	87	85	84

^aR(-) represented RF-negative RA patients
^bA(-) represented ACPA-negative RA patients
^cR&A(-) represented RF&ACPA-negative RA patients
^dThe prediction accuracy refers to the ratio of the number of samples correctly predicted as RA to the total number of samples actually diagnosed as RA in corresponding subsets using the regular diagnostic methods

verification^{33, 34}, all the 21 biomarkers and its corresponding cut off points were applied to a validation set consisted of 187 RA patients and 84 healthy samples. This independent, blinded analysis confirmed the performance of the biomarkers established in training set, as demonstrated by an overall AUC of 0.804, sensitivity of 82% and specificity of 79%. Notably, the prediction capacity of these biomarkers for RF- and/or ACPA-negative RA patients were verified in a much bigger subset of the validation cohort ($n=97$ for RF-negative patients, $n=55$ for ACPA-negative patients, and $n=45$ for RF&ACPA-negative patients), with an overall prediction accuracy of 81%, 80% and 80%, respectively. Furthermore, performance of SGm1, SGm2 and their combination were validated by their comparable AUC, sensitivity, specificity, and prediction accuracy for RF and/or ACPA-negative patients in the validation set as that in training set (Tables 1, 2).

In addition, although some differences of clinical features between the training set and the validation set were observed, the values in both data sets fell into the range of clinical settings, thus reflected the clinical heterogeneity of RA patients. Validation on such heterogeneous data sets just allowed for the evaluation of the generalizability of the established model to wider RA populations.

As RF-independent biomarkers, SGm1 and SGm2 may be used for the diagnosis of autoantibody-negative RA patients. To this end, we further examined the specificity of SGm1 and SGm2 for RF/ACPA-negative RA patients by analyzing patients of another RF/ACPA-negative inflammatory disorder, ankylosing

spondylitis (AS, $n=34$), as well as a degenerative joint disease osteoarthritis (OA, $n=26$). As shown in Fig. 7c–e, SGm1, which showed significantly enhanced relative abundance in both RF-positive and -negative RA patients, displayed almost unchanged levels in AS and OA patients as compared to their age and gender-matched healthy controls. Of interest, SGm2 exhibited disease-associated changes in RA, OA and AS (Fig. 7c–e). It showed significantly elevated levels in RA patients, obvious reduced levels in OA patients, and no significant changes in AS patients, as compared to age and gender-matched healthy controls of each. These findings confirmed the specificity of SGm1 and SGm2 upregulation for RA patients, regardless they are RF/ACPA-positive or negative.

Compared with previously reported serum N-glycan biomarkers, such as the ratio of G₀/G₁ (Fig. 7c–e)^{3–5, 10, 35}, the newly identified N-glycan biomarkers, especially SGm1 and SGm2, exhibited higher potential for the classification of RF/ACPA-negative RA patients, (Table 2). Of note, comparing to other disease controls, no specific change was observed for G₀/G₁, as indicated by the observation of significant increased levels in both AS and OA as compared to respective controls, showing that this biomarker was not specific for RA patients.

Discussion

Any changes in the structures or levels of even trace N-glycans can result in significant physiological/pathological events. By sharply increasing the glycome coverage and depth, our

Table 3 Clinical features of RA patients and healthy individuals

Clinical features	Training set			Validation set		
	RA patients (n = 90)	Healthy individuals (n = 57)	P-value ^a	RA patients (n = 187)	Healthy individuals (n = 84)	P-value ^a
Age, years	57.2 ± 13.6	56.7 ± 14.2	0.85	55.8 ± 12.7	52.9 ± 10.8	0.07
Gender (female), %	84.4	71.9	0.37	75.9	67.9	0.18
Disease duration, years	7.4 ± 7.1	ND		7.3 ± 6.8	ND	
<i>Serology parameters:</i>						
RF positive, %	78.9	ND		48.1 ^b	ND	
RF, IU ml ⁻¹	410.3 ± 695.5	ND		376.6 ± 672.8	ND	
ACPA positive, %	83.3	ND		70.6	ND	
ACPA, units per ml	717.6 ± 698.5	ND		840.5 ± 1115.4	ND	
ESR, mm h ⁻¹	48.6 ± 31.8	ND		42.3 ± 29.1	ND	
CRP, mg l ⁻¹	2.5 ± 3.2	ND		10.9 ± 21.1	ND	
<i>Medication used:</i>						
Leflunomide, %	72.2	ND		58.8	ND	
Methotrexate, %	51.1	ND		47.1	ND	
Hydroxychloroquine, %	24.4	ND		21.9	ND	

ACPA anti-citrullinated protein antibodies, CRP C-reactive protein, ESR erythrocyte sedimentation rate, ND not determined, RF rheumatoid factor

Data are shown as mean ± s.d.

^aP-values were calculated by the non-parametric Mann-Whitney test for age and by the Fisher's exact test for gender, two-sided

^bRF-negative patients were collected on purpose to confirm the prediction capacity of the biomarkers for RF-negative samples, hence the RF-positive percentage in validation set was much lower than that in training set

chip-based approach provides an early glimpse into the remarkable structural complexity of N-glycans resulting from microheterogeneity expressions, such as sulfation, phosphorylation and acetylation. Moreover, as all N-glycans share a common core sugar sequence, the TiO₂-PGC chip-based glycomic approach is applicable for profiling N-glycans released from any single glycoprotein or total glycoproteins. N-glycosylation occurs on numerous secreted and membrane-bound glycoproteins¹², and N-glycan components are often the crucial functional determinants of biological events. Therefore, our glycomic approach will rapidly position itself as one of the most important tools for addressing certain key biological and pathological questions. Furthermore, because of the conserved biosynthesis of N-glycans across metazoa, plants, yeast and even bacteria¹², this on-chip glycomic approach can be further extrapolated to the quality control of antibody-based drugs¹ and design of vaccines because antigen glycosylation, including N-glycosylation, has been increasingly appreciated as essential in adaptive immune activation³⁶.

Our study represents the first glycomic method that enables the detection of low-abundance and even trace novel sulfated N-glycans on IgGs. Because of the conserved biosynthesis of N-glycans, the novel sulfated N-glycans are most likely present on other glycoproteins. It was believed that sulfation of N-glycans can significantly alter biological recognition and/or facilitate rapid clearance of the protein from the body^{37–40}. Especially, N-glycans with sulfo-modifications constitute important recognition codes in cell adhesion, e.g., a glucosamine containing sulfate group in position O-6 and an N-acetyl group was the preferred epitope for the immune recognition^{31, 41–43}. The effect of sulfated N-glycans on IgG is yet to be known. However, as terminal sugar of N-glycan may affect the function of IgG dramatically via altering the structure of C_γ2 domain, we speculated that sulfated N-glycans on IgG may induce significant structural alterations in C_γ2 domain, and subsequently change the ligand specificity and biological functions of IgG, which need to be studied in the future.

Next to the early-appreciated correlation between hypogalactosylation of serum total IgGs and the severity of RA, more attention have been made on the analysis of specific glycan profile

of ACPA antibodies that is distinct from the profile of total serum IgGs^{44, 45}. Furthermore, increasing experimental evidence points to a critical role of autoantibody-associated glycosylation of IgGs in RA. For instance, extensive glycosylation of ACPA-IgG variable domains were found to modulate binding to citrullinated antigens in RA³, while ACPA antibody itself were demonstrated to acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of RA⁴⁶. In addition, agalactosyl glycoforms of IgG autoantibodies were suggested to be pathogenic for RA^{35, 47}. Despite these findings, specific glycosylation-based biomarkers with diagnostic capabilities for autoantibody-negative RA are still absent⁴⁸. In this context, the trace N-glycan biomarkers identified in our study, especially SGM1 and SGM2, could have important clinical implications for the diagnosis of RA as they are RF/ACPA-independent and RA-specific. In addition, as N-glycan biomarkers of total IgG rather than autoantibody-specific IgG, the on-chip method readily lends itself to clinical applications. Even though, for the discrimination of RA from other inflammatory arthritis which is a crucial clinical question^{49–51}, further studies need to be performed to address the specificity of SGM1 and SGM2.

In summary, we have developed an innovative platform based on TiO₂-PGC chip for in-depth glycomic research of human serum IgGs. This innovative approach facilitated the identification of a high number of glycans on IgGs and the discovery of a number of sulfated glycans on IgGs for the first time. We have also demonstrated the significance of this cutting-edge glycomic platform in discovering potential biomarkers of RA disease. Two sulfated glycans have been proved to exhibit notably high capacity for the classification of both rheumatoid factor (RF)-positive and negative RA patients, as well as anti-citrullinated protein antibodies (ACPA)-positive and negative RA patients, showing notable and specific value of N-glycan as biomarkers for RA. Application of this novel approach in a wider range of glycomic research may reveal potential N-glycosylation-associated biomarkers for other autoimmune and infectious diseases. Further studies on the biosynthesis and biological functions of novel sulfated glycans identified in this study may contribute potentially to the pathological study of autoimmune diseases.

Methods

Patients and collection of serum samples. A total of 141 healthy individuals and 277 RA patient samples from three independent set were analyzed in this study, comprising a training set and validation set. The clinical features of RA patients and healthy individuals were shown in Table 3. The training set consisted of 57 healthy individuals and 90 RA patient samples (including a small subset of RF-negative patients ($n = 19$) and subset of ACPA-negative patients ($n = 15$)). The samples were collected from the Division of Rheumatology of Jiujiang No. 1 People's Hospital (Jiujiang, China). The validation set consisted of 84 healthy individuals and 187 RA patient samples (including a bigger subset of RF-negative patients ($n = 97$) and subset of ACPA-negative patients ($n = 55$)). Samples were obtained from Jiujiang and the Department of Rheumatology and Immunology of Peking University People's Hospital (Beijing, China).

All RA patients, including RF-negative and ACPA-negative RA patients, fulfilled the 2010 ACR/EULAR Classification Criteria for Rheumatoid Arthritis. Firstly, all of the patients must meet two requirements: (1) patient with at least 1 joint with definite clinical synovitis (swelling) and (2) the synovitis is not better explained by "another" disease. With the fulfillment of the two requirements, patients with a total score of 6 or greater from the individual scores in 4 domains were classified as definite RA, whereas patients with erosions typical for RA were classified as RA even without application or fulfillment of the scoring system of 2010 ACR/EULAR Classification Criteria for RA⁵². Samples from patients of other RF-negative chronic inflammatory disorders, ankylosing spondylitis (AS, $n = 34$) and osteoarthritis (OA, $n = 26$) and controls of each ($n = 27$ for AS; $n = 45$ for OA) were collected from the Division of Rheumatology of Jiujiang No. 1 People's Hospital and Guangdong Provincial Hospital of Chinese Medicine (Guangzhou, China). AS patients were classified according to the modified New York Criteria⁵³ and European Spondyloarthropathy Study Group Criteria⁵⁴. All knee OA patients fulfilled the American College of Rheumatology (ACR) classification criteria⁵⁵. The clinical features of AS and OA patients were shown in Supplementary Table 4 and 5.

Serum samples were retrieved from hospital following the same protocol. All sera were stored at -80°C prior to use. The study was approved by the Ethic Committees of the relevant hospitals and informed consent was obtained from all patients.

Capture of IgGs from serum samples. IgGs were isolated using Protein A. Briefly, rProtein A Sepharose 4 Fast Flow beads were applied to a 96-well filter plate at 50 μl per well. After washing twice with five volumes of binding buffer (20 mM sodium phosphate, pH 7.0), 250 μl binding buffer and 10 μl serum were successively applied to each well. The plate was sealed and incubated on a shaker at room temperature for 15 min. The filtrate was collected in a V-bottom collection plate by centrifugation (1000 r.p.m., 5 min). The retained beads were washed twice with 250 μl binding buffer. IgGs were then eluted twice with 200 μl elution buffer (0.1 M glycine buffer, pH 2.7) into a new V-bottom collection plate. Next, 30 μl neutralizing buffer (1 M Tris-HCl, pH 9.0) was added for neutralization. The obtained IgG samples were then transferred to 30 K centrifuge filter units for buffer exchange, and the resulting water solutions were concentrated to a final volume of 30 μl . The amount of captured IgGs in each sample was quantitated using a Bio-Rad protein assay. The purity of the captured IgGs was examined using SDS-PAGE and HPLC.

Release of N-glycans. N-glycans were cleaved by using PNGase F. 50 μg IgGs from each patient sample was diluted with 100 mM ammonium bicarbonate buffer (pH 7.4) to a final concentration of 1 μg μl^{-1} . Then, 0.5 μl PNGase F was added, followed by a 16 h incubation at 37°C . The cleaved N-glycans were directly loaded onto the preconditioned C₁₈ cartridge and washed with 1.0 ml distilled water to remove the de-glycosylated protein. The flow-through and water eluates were combined and dried by speed vacuum. The dried residues were reconstituted in 100 μl distilled water and stored at -80°C before analysis.

Design and fabrication of TiO₂-PGC Chip. A customized TiO₂-PGC chip composed of a sandwich-like enrichment column, which included a TiO₂ column between two PGC trapping column, was fabricated. The customized TiO₂-PGC chip consists of a 75 $\mu\text{m} \times 150$ mm PGC analytical column (PGC, 5 μm) and a three-sectioned enrichment column, including a 100 nl PGC trapping column (PGC 5 μm), a 45 nl TiO₂ column, and a second 100 nl PGC trapping column (Agilent, Waldbronn, Germany). The sandwich design of the enrichment column allowed for the enrichment of acidic N-glycans on TiO₂, whereas neutral N-glycans were eluted and analyzed in the first step. The acidic N-glycans were subsequently eluted by injecting an elution buffer (0.5% ammonia solution in water) in the second step. The acidic N-glycan binding capacity of the TiO₂-PGC chip was evaluated using a breakthrough experiment. The online enrichment performance of the TiO₂-PGC chip for acidic N-glycans was compared with the offline SAX method.

HPLC chip/MS. An Agilent 1260 Infinity HPLC Chip LC system was coupled to an Agilent 6550 iFunnel Quadrupole Time-of-Flight (Q-TOF) mass spectrometer (MS) for N-glycan profiling or coupled to an Agilent 6490 iFunnel Triple

Quadrupole (QQQ) MS for N-glycan quantification. The Agilent 1260 Infinity HPLC Chip system was equipped with a HiP micro ALS sampler with a 40 μl sample loop, a nanoflow pump, a capillary pump, a HPLC Chip Cube Interface, a thermostat and a μ -degasser. A 25 μm ID PEEK capillary was used for sample transfer to prevent the dissolution of fused silica by the high-pH elution buffer.

On-chip enrichment of acidic N-glycans. The TiO₂-PGC chip was operated in the forward flush mode. First, 2 μl of the sample were injected and transferred onto the enrichment column using 0.6% acetic acid, 2% FA and 2% ACN in water at a flow rate of 3 μl min^{-1} . The chip valve was switched 2 min after injection to place the enrichment column in-line with the analytical column. The mobile phase used in the nanoflow pump was optimized for neutral N-glycans and consisted of 1% FA in water (A) and ACN (B). The gradient was performed at a flow rate of 0.5 μl min^{-1} as follows: 5% B for 6 min, 5–60% B over 10 min, and 80% B for 3 min. The acidic N-glycans were subsequently eluted by injecting 5 μl elution buffer (0.5% ammonia solution in water). The analysis of the enriched acidic N-glycans was performed by switching the enrichment column in-line with the analytical column 1 min after injection. The mobile phase optimized for the acidic N-glycans was used. Mobile phase A was 0.5% FA in water adjusted to pH 3 with the ammonia solution, and mobile phase B was 1% FA in ACN. The flow rate was 0.5 μl min^{-1} , and the gradient was as follows: 5% B for 1 min, 5–60% B over 10 min, and 80% B for 3 min. An equilibrium time of 18 min was set before each injection.

N-glycan profiling by HPLC chip-Q-TOF MS. The structures of the N-glycans were characterized based on high-resolution MS and MS/MS data obtained on Q-TOF MS in the positive mode. The dry-gas (N₂) temperature and flow rate were 225 $^{\circ}\text{C}$ and 11 l min^{-1} , respectively. The MS spectra were acquired in the positive mode, and the mass range was 500–3000 m/z with an acquisition time of 1 spectrum per sec. Mass correction was enabled using reference masses of 922.0098 m/z and 1221.9906 m/z . The mass range of the MS/MS experiments was 100 to 3000 m/z . The spectra were acquired in the targeted MS/MS mode with an MS acquisition rate of 2 spectra per sec and an MS/MS acquisition rate of 3 spectra per sec. The collision energy (CE) was set to 10–40 eV.

Quantitative profiling of N-glycans on HPLC Chip-QQQ-MS. The released N-glycans were quantitatively profiled on Chip-QQQ-MS using MRM method in positive mode. The dry-gas (N₂) temperature and flow rate were 225 $^{\circ}\text{C}$ and 11 l min^{-1} , respectively. The RF voltage amplitude of the high-pressure and low-pressure ion funnels were 150 and 200 V, respectively. The dwell time was set as 10–50 ms. All of the raw data were processed using Agilent MassHunter Qualitative Analysis B.06.00 and Agilent MassHunter Quantitative Analysis B.06.00 software. The optimized method was validated for linearity, sensitivity, recovery and repeatability. The abundance of each N-glycan composition was relatively quantified based on the peak areas of their MRM chromatograms and expressed as a percentage of summed peak areas for total N-glycans within each IgG samples.

Identification of sulfation sites using exoglycosidases. Four exoglycosidases, Sialidase C, β 1–4 Galactosidase, β -N-Acetyl glucosaminidase, and α 1–2,3 Mannosidase, were employed to hydrolyze the N-glycans and determine the monosaccharides linked to the sulfate groups.

Sialic acids were released by enzymatic digestion using Sialidase C. Briefly, N-glycans from 20 μg IgGs were reconstituted with 100 μl 50 mM NH₄Ac (pH 5.0), and 5 μl Sialidase C¹⁷ (0.05 units) was added subsequently. The solution was incubated at 37°C for 18 h, and the digestion was then terminated by heating the solution in boiling water for 5 min. The digestion was evaporated by speed vacuum and then resuspended into 40 μl H₂O. After centrifugation at 14,000 \times g for 15 min, 30 μl of the supernatant was loaded into a vial insert with 1 μl of the acidic N-glycan IS. For the blank samples, 5 μl H₂O was added instead.

β 1–4 Galactosidase was employed to digest β 1–4-linked lactose. In a total reaction volume of 10 μl , N-glycans from 20 μg IgGs and 1 μl β 1–4 Galactosidase (8 units) were incubated in a sodium citrate (50 mM, pH 6.0) and NaCl (100 mM) reaction buffer for 1 h at 37°C . After being diluted to 40 μl , the solution was centrifuged at 14,000 \times g for 15 min, and 30 μl of the supernatant was loaded into a vial insert with 1 μl of the acidic N-glycan IS. For the blank samples, 1 μl H₂O was used instead of β 1–4 Galactosidase.

β -N-Acetyl glucosaminidase was used to cleave the β -N-Acetyl glucosamine residues from oligosaccharides. Briefly, N-glycans from 20 μg IgGs and 1 μl β -N-Acetyl glucosaminidase (4 units) were incubated in a sodium citrate BSA buffer (50 mM, pH 6.0) for 4 h at 37°C in a total volume of 10 μl . After being diluted to 40 μl , the solution was centrifuged at 14,000 \times g for 15 min, and 30 μl of the supernatant was loaded into a vial insert with 1 μl of the acidic N-glycan IS. For the blank samples, 1 μl H₂O replaced β -N-Acetyl glucosaminidase.

α 1–2,3 Mannosidase was employed to digest the α 1–2,3 mannose residues from oligosaccharides. Briefly, N-glycans from 20 μg IgGs and 1 μl α 1–2,3 Mannosidase (32 units) were incubated in a sodium acetate (50 mM, pH 5.5) and CaCl₂ (5 mM) BSA buffer for 1 h at 37°C in a total volume of 10 μl . After being diluted to 40 μl , the solution was centrifuged at 14,000 \times g for 15 min, and 30 μl of the supernatant

was loaded into a vial insert with 1 μ l of the acidic N-glycan IS. For the blank samples, 1 μ l H₂O replaced α 1–2,3 Mannosidase.

All of the samples in the exoglycosidase experiments were prepared in duplicate.

Statistical analysis. Comparisons were performed with the Mann–Whitney test for non-parametric continuous variables and Fisher's exact test for categorical variables. Continuous variables were presented as mean \pm s.d. Statistical significances were calculated using GraphPad Prism 6 (GraphPad Software, La Jolla, CA, USA). A two-sided *P*-value of < 0.05 was considered statistically significant. The sample size was not predetermined by statistical method. No data points were excluded from the analysis. Support Vector Machine (SVM) approach in Weka (University of Waikato), the software widely used for data mining, was used to generate the classification model. Attribute selection was performed using correlation-based feature subset selection (CfsSubsetEval) to select value subsets that correlate highly with the class value and lowly with each other. The best-first search strategy was used to navigate attribute subsets in search space. The quality or performance of the predicted models were evaluated by using the receiver operating characteristic (ROC) curve, which is obtained by calculating the sensitivity and specificity of the test at every possible cut off point, and plotting sensitivity (the proportion of true positive results) against 1-specificity (the proportion of false positive results). The method of Youden index (*J*) was employed to identify optimal cutoff points based on sensitivity, specificity and the ROC curve⁵⁶. *J* was defined as the maximum vertical distance between the ROC curve and the diagonal or chance line and was calculated as $J = \text{maximum}\{\text{sensitivity} + \text{specificity} - 1\}$. Using this measure, the cut off point on the ROC curve which corresponded to *J* at which (sensitivity + specificity – 1) was maximized, was taken to be the optimal cut off point. The prediction model was then measured for sensitivity by using Eq. (1) and specificity by using Eq. (2),

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}), \quad (1)$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}), \quad (2)$$

where true positives (TP) denote the correct classifications of positive examples, true negatives (TN) are the correct classification of negative examples, false positives (FP) denote the incorrect classification of negative examples into the positive class and false negatives (FN) represent the incorrect classification of positive examples into the negative class.

The area under the ROC curve (AUC) was a reflection of how good the biomarker was at distinguishing RA patients and healthy individuals. The AUC served as a single measure, independent of prevalence, that summarized the discriminative ability of a test across the full range of cutoffs. The closer the AUC to 1, the better the overall diagnostic performance of the biomarker, and the closer it to 0.5, the poorer the test. All individual potential N-glycan biomarkers were further paired as different combinations and analyzed by the same strategy. Moreover, all potential single N-glycan biomarkers and combinations were applied in the prediction of RF-negative patients ($n = 19$) and ACPA-negative patients ($n = 15$), the prediction accuracy (%) was calculated as ((total number – mispredicted number) / total number \times 100%).

For verification, the same classification models based on the identified individual or combined N-glycan biomarkers, together with their corresponding optimal cut off points generated from the training data sets, were applied to a validation set (187 RA patients and 84 healthy individuals). The prediction accuracy of RF-negative patients ($n = 97$) and ACPA-negative patients ($n = 55$) in the validation set were calculated by the same methods as that used in the training set.

Data availability. The data that support the findings of the study are available from the corresponding author upon reasonable request.

Received: 12 October 2015 Accepted: 19 July 2017

Published online: 20 September 2017

References

- Jefferis, R. Glycosylation as a strategy to improve antibody-based therapeutics. *Nat. Rev. Drug Discov.* **8**, 226–234 (2009).
- Maverakis, E. et al. Glycans in the immune system and the altered glycan theory of autoimmunity: a critical review. *J. Autoimmun.* **57**, 1–13 (2015).
- Rombouts, Y. et al. Extensive glycosylation of ACPA-IgG variable domains modulates binding to citrullinated antigens in rheumatoid arthritis. *Ann. Rheum. Dis.* **75**, 578–585 (2016).
- Parekh, R. B. et al. Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **316**, 452–457 (1985).
- Parekh, R. B. et al. Galactosylation of IgG associated oligosaccharides: reduction in patients with adult and juvenile onset rheumatoid arthritis and relation to disease activity. *Lancet* **1**, 966–969 (1988).
- Wang, J. R. et al. Glycomic signatures on serum IgGs for prediction of postvaccination response. *Sci. Rep.* **5**, 7648 (2015).
- An, H. J., Kronewitter, S. R., de Leoz, M. L. & Lebrilla, C. B. Glycomics and disease markers. *Curr. Opin. Chem. Biol.* **13**, 601–607 (2009).
- Shinzaki, S. et al. IgG oligosaccharide alterations are a novel diagnostic marker for disease activity and the clinical course of inflammatory bowel disease. *Am. J. Gastroenterol.* **103**, 1173–1181 (2008).
- Schwab, I. & Nimmerjahn, F. Intravenous immunoglobulin therapy: how does IgG modulate the immune system? *Nat. Rev. Immunol.* **13**, 176–189 (2013).
- Harre, U. et al. Glycosylation of immunoglobulin G determines osteoclast differentiation and bone loss. *Nat. Commun.* **6**, 6651 (2015).
- Alley, W. R. Jr, Mann, B. F. & Novotny, M. V. High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem. Rev.* **113**, 2668–2732 (2013).
- Varki, A. et al. *Essentials of Glycobiology* 2nd edn (Cold Spring Harbor Laboratory Press, 2009).
- Go, E. P. et al. Characterization of glycosylation profiles of HIV-1 transmitted/founder envelopes by mass spectrometry. *J. Virol.* **85**, 8270–8284 (2011).
- Ruhaak, L. R., Miyamoto, S. & Lebrilla, C. B. Developments in the identification of glycan biomarkers for the detection of cancer. *Mol. Cell. Proteomics* **12**, 846–855 (2013).
- Gornik, O., Pavic, T. & Lauc, G. Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. *Biochim. Biophys. Acta* **1820**, 1318–1326 (2012).
- Kaneko, Y., Nimmerjahn, F. & Ravetch, J. V. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* **313**, 670–673 (2006).
- Wu, Z. L., Prather, B., Ethen, C. M., Kalyuzhny, A. & Jiang, W. Detection of specific glycosaminoglycans and glycan epitopes by in vitro sulfation using recombinant sulfotransferases. *Glycobiology* **21**, 625–633 (2011).
- Toyoda, M., Narimatsu, H. & Kameyama, A. Enrichment method of sulfated glycopeptides by a sulfate emerging and ion exchange chromatography. *Anal. Chem.* **81**, 6140–6147 (2009).
- Sondermann, P., Pincetic, A., Maamary, J., Lammens, K. & Ravetch, J. V. General mechanism for modulating immunoglobulin effector function. *Proc. Natl Acad. Sci. USA* **110**, 9868–9872 (2013).
- Thomsson, K. A., Backstrom, M., Holmen Larsson, J. M., Hansson, G. C. & Karlsson, H. Enhanced detection of sialylated and sulfated glycans with negative ion mode nanoliquid chromatography/mass spectrometry at high pH. *Anal. Chem.* **82**, 1470–1477 (2010).
- Zhang, Q. et al. Methylamidation for isomeric profiling of sialylated glycans by nanoLC-MS. *Anal. Chem.* **86**, 7913–7919 (2014).
- Mohammed, S. et al. Chip-based enrichment and NanoLC-MS/MS analysis of phosphopeptides from whole lysates. *J. Proteome Res.* **7**, 1565–1571 (2008).
- Rajh, T., Dimitrijevic, N. M., Bissonnette, M., Koritarov, T. & Konda, V. Titanium dioxide in the service of the biomedical revolution. *Chem. Rev.* **114**, 10177–10216 (2014).
- Palmisano, G. et al. A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development. *Mol. Cell. Proteomics* **11**, 1191–1202 (2012).
- Engholm-Keller, K. & Larsen, M. R. Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds--applications in acidic modification-specific proteomics. *J. Proteomics* **75**, 317–328 (2011).
- Palmisano, G. et al. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **5**, 1974–1982 (2010).
- Kronewitter, S. R. et al. The development of retrosynthetic glycan libraries to profile and classify the human serum N-linked glycome. *Proteomics* **9**, 2986–2994 (2009).
- Bakovic, M. P. et al. High-throughput IgG Fc N-glycosylation profiling by mass spectrometry of glycopeptides. *J. Proteome Res.* **12**, 821–831 (2013).
- Pfeiffer, G. et al. Structure elucidation of sulphated oligosaccharides from recombinant human tissue plasminogen activator expressed in mouse epithelial cells. *Glycobiology* **2**, 411–418 (1992).
- van Rooijen, J. J., Kamerling, J. P. & Vliegthart, J. F. Sulfated di-, tri- and tetraantennary N-glycans in human Tamm-Horsfall glycoprotein. *Eur. J. Biochem.* **256**, 471–487 (1998).
- Kamerling, J. P., Rijkse, I., Maas, A. A., van Kuik, J. A. & Vliegthart, J. F. Sulfated N-linked carbohydrate chains in porcine thyroglobulin. *FEBS Lett* **241**, 246–250 (1988).
- de Waard, P., Koorevaar, A., Kamerling, J. P. & Vliegthart, J. F. Structure determination by 1H NMR spectroscopy of (sulfated) sialylated N-linked

- carbohydrate chains released from porcine thyroglobulin by peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase-F. *J. Biol. Chem.* **266**, 4237–4243 (1991).
33. Rassi, A. Jr. et al. Development and validation of a risk score for predicting death in Chagas' heart disease. *N. Engl. Med.* **355**, 799–808 (2006).
 34. Agesen, T. H. et al. ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. *Gut* **61**, 1560–1567 (2012).
 35. Tomana, M., Schrohenloher, R. E., Koopman, W. J., Alarcon, G. S. & Paul, W. A. Abnormal glycosylation of serum IgG from patients with chronic inflammatory diseases. *Arthritis Rheum.* **31**, 333–338 (1988).
 36. Wolfert, M. A. & Boons, G.-J. Adaptive immune activation: glycosylation does matter. *Nat. Chem. Biol.* **9**, 776–784 (2013).
 37. Fiete, D., Srivastava, V., Hindsgaul, O. & Baenziger, J. U. A hepatic reticuloendothelial cell receptor specific for SO4-4GalNAc beta 1,4GlcNAc beta 1,2Man alpha that mediates rapid clearance of lutropin. *Cell* **67**, 1103–1110 (1991).
 38. Kawashima, H. Roles of sulfated glycans in lymphocyte homing. *Biol. Pharm. Bull.* **29**, 2343–2349 (2006).
 39. Taguchi, T. et al. Occurrence and structural analysis of highly sulfated multiantennary N-linked glycan chains derived from a fertilization-associated carbohydrate-rich glycoprotein in unfertilized eggs of *Tribolodon hakonensis*. *Eur. J. Biochem.* **238**, 357–367 (1996).
 40. Green, E. D., Morishima, C., Boime, I. & Baenziger, J. U. Structural requirements for sulfation of asparagine-linked oligosaccharides of lutropin. *Proc. Natl Acad. Sci. USA* **82**, 7850–7854 (1985).
 41. Bai, X., Brown, J. R., Varki, A. & Esko, J. D. Enhanced 3-O-sulfation of galactose in Asn-linked glycans and Maackia amurensis lectin binding in a new Chinese hamster ovary cell line. *Glycobiology* **11**, 621–632 (2001).
 42. Mitoma, J. et al. Critical functions of N-glycans in L-selectin-mediated lymphocyte homing and recruitment. *Nat. Immunol.* **8**, 409–418 (2007).
 43. Couto, A. S. et al. An anionic synthetic sugar containing 6-SO₃-NACglc mimics the sulfated cruzipain epitope that plays a central role in immune recognition. *FEBS J.* **279**, 3665–3679 (2012).
 44. Scherer, H. U. et al. Glycan profiling of anti-citrullinated protein antibodies isolated from human serum and synovial fluid. *Arthritis Rheum.* **62**, 1620–1629 (2010).
 45. Scherer, H. U. et al. Immunoglobulin 1 (IgG1) Fc-glycosylation profiling of anti-citrullinated peptide antibodies from human serum. *Proteomics Clin. Appl.* **3**, 106–115 (2009).
 46. Rombouts, Y. et al. Anti-citrullinated protein antibodies acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of rheumatoid arthritis. *Ann. Rheum. Dis.* **74**, 234–241 (2015).
 47. Rademacher, T. W., Williams, P. & Dwek, R. A. Agalactosyl glycoforms of IgG autoantibodies are pathogenic. *Proc. Natl Acad. Sci. USA* **91**, 6123–6127 (1994).
 48. Sun, J., Zhang, Y., Liu, L. & Liu, G. Diagnostic accuracy of combined tests of anti cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis: a meta-analysis. *Clin. Exp. Rheumatol.* **32**, 11–21 (2014).
 49. Kim, S. et al. Global metabolite profiling of synovial fluid for the specific diagnosis of rheumatoid arthritis from other inflammatory arthritis. *PLoS ONE* **9**, e97501 (2014).
 50. Krenn, V. et al. Grading of chronic synovitis—a histopathological grading system for molecular and diagnostic pathology. *Pathol. Res. Pract.* **198**, 317–325 (2002).
 51. Baillet, A. et al. Synovial fluid proteomic fingerprint: S100A8, S100A9 and S100A12 proteins discriminate rheumatoid arthritis from other inflammatory joint diseases. *Rheumatology* **49**, 671–682 (2010).
 52. Aletaha, D. et al. 2010 rheumatoid arthritis classification criteria: an American college of rheumatology/European league against rheumatism collaborative initiative. *Arthritis Rheum.* **62**, 2569–2581 (2010).
 53. van der Linden, S., Valkenburg, H. A. & Cats, A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum.* **27**, 361–368 (1984).
 54. Dougados, M. et al. The European spondylarthropathy study group preliminary criteria for the classification of spondylarthropathy. *Arthritis Rheum.* **34**, 1218–1227 (1991).
 55. Altman, R. et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. *Arthritis Rheum* **29**, 1039–1049 (1986).
 56. Akobeng, A. K. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr.* **96**, 644–647 (2007).

Acknowledgements

We thank all members of the group who contributed to development of this novel approach over the years along with government sponsor: Macao Science and Technology Development Fund (Project code: 020/2013/A1 to J.Z.H. and 023/2016/AFJ to W.J.R.). We further acknowledge the technical support from engineers from Agilent Technologies: Jimmy Chan, Winnie Huang and Kelvin Wong.

Author contributions

Project design: Z.-H.J., L.L., S.J. and J.-R.W. Chip design: R.G., Z.-H.J., J.-R.W. and S.L. Mass spectrometry experiment and data analysis: W.-N.G., L.-F.Y., H.H., Q.M., T.-T.T. and J.R.W. Statistics: Y.L. and H.H.H. Clinical sample: J.L., M.J., H.Y., Z.-G.L. and X.Z. Writing: J.-R.W., Z.-H.J., S.J. and L.L.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-00662-w.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017