Anal Bioanal Chem (2014) 406:7925–7935
DOI 10.1007/s00216-014-8261-2

RESEARCH PAPER

Rapid-throughput glycomics applied to human milk oligosaccharide profiling for large human studies

Sarah M. Totten · Lauren D. Wu · Evan A. Parker · Jasmine C. C. Davis · Serenus Hua · Carol Stroble · L. Renee Ruhaak · Jennifer T. Smilowitz · J. Bruce German · Carlito B. Lebrilla

Received: 26 August 2014 / Revised: 2 October 2014 / Accepted: 8 October 2014 / Published online: 31 October 2014

Abstract Glycomic analysis is the comprehensive determination of glycan (oligosaccharide) structures with quantitative information in a biological sample. Rapid-throughput glycomics is complicated due to the lack of a template, which has greatly facilitated analysis in the field of proteomics. Furthermore, the large similarities in structures make fragmentation spectra (as obtained in electron impact ionization and tandem mass spectrometry) less definitive for identification as it has been in metabolomics. In this study, we develop a concept of rapid-throughput glycomics on human milk oligosaccharides, which have proven to be an important bioactive component of breast milk, providing the infant with protection against pathogenic infection and supporting the establishment of a healthy microbiota. To better understand the relationship between diverse oligosaccharides structures and their biological function as anti-pathogenic and prebiotic compounds, large human studies are needed, which necessitate rapid- to high-throughput analytical platforms. Herein, a complete glycomics methodology is presented, evaluating the most effective human milk oligosaccharide (HMO) extraction protocols, the linearity and reproducibility of the nano-liquid chromatography chip time-of-flight mass spectrometry (nano-LC chip-TOF MS) method, and the efficacy of newly developed, in-house software for chromatographic peak alignment that allows for rapid data analysis. High instrument stability and retention time reproducibility, together with the successful automated alignment of hundreds of features in hundreds of milk samples, allow for the use of an HMO library for rapid assignment of fully annotated structures.

Keywords Human milk oligosaccharides · Rapid-throughput glycomics · Nano-flow liquid chromatography-mass spectrometry

Introduction

Human milk oligosaccharides (HMOs) are an abundant component of human milk and have been linked to several aspects of infant health. Studies have shown that specific oligosaccharide structures protect breast-fed infants from infection by inhibiting the binding of pathogens to host cell ligands and by serving as prebiotics for commensal bacteria in the gut [1–5]. Detailed structural information and quantitation are key to better understand the biological function of HMOs. High-throughput analytical platforms are needed to rapidly assign specific structure to HMOs in large groups of individuals required in human studies. Although there has been progress towards developing the analytical techniques needed to comprehensively study the role of glycosylation in disease and nutrition, increasing the throughput in glycomics studies continues to present challenges due to inherent glycan heterogeneity and structural diversity [6]. HMOs exemplify this point, with several hundreds of unconjugated oligosaccharide

Electronic supplementary material The online version of this article (doi:10.1007/s00216-014-8261-2) contains supplementary material, which is available to authorized users.

S. M. Totten · L. D. Wu · E. A. Parker · J. C. C. Davis · S. Hua · C. Stroble · L. R. Ruhaak · C. B. Lebrilla (✉)
Department of Chemistry, University of California, Davis, CA 95616, USA
e-mail: cblebrilla@ucdavis.edu

S. Hua
Asia Glycomics Reference Site, Chungnam National University, Daejeon 305-764, South Korea

J. T. Smilowitz · J. B. German
Foods for Health Institute, University of California, Davis, CA 95616, USA

J. T. Smilowitz · J. B. German
Department of Food Science & Technology, University of California, Davis, CA 95616, USA

structures native to the breast milk glycome, many of which belong to a family of structural isomers, making separation and identification a challenging endeavor. Currently, a variety of liquid chromatography (LC) methods with spectroscopic detection are employed for oligosaccharide analysis, including reverse-phase HPLC of derivatized glycans with spectroscopic detection, hydrophilic interaction liquid chromatography (HILIC) of glycans with a fluorophore label, and high-pH anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) [2, 7–10]. These methods are not ideal for rapidly obtaining full structural characterization of the hundreds of native structures due to the lack of commercially available standards and the inability to resolve overlapping peaks from co-eluting compounds and thus require additional off-line analyses for complete elucidation. However, when coupled to mass spectrometry, typically nano-ESI-MS (electrospray ionization-mass spectrometry), co-eluting HMO species can be readily separated based on mass, allowing for higher compound resolution and more accurate quantitation [11–15]. In addition to liquid chromatography-based methods, a recent study by Kottler et al. describes the use of multiplexed capillary gel electrophoresis with laser-induced fluorescence as a high-throughput method for glycoanalysis, demonstrating the ability to identify and quantify approximately 17 oligosaccharide structures based on "fingerprint" electropherograms [16]. Also in a recent study, Blank et al. describe the use of mass fingerprinting using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) for a high-throughput assignment of Lewis blood groups present on milk glycans [17].

Even with advances in the analytical techniques used for glycan analysis, rapid-throughput structure elucidation of carbohydrates in complex biological mixtures is a difficult task, but can be greatly facilitated by the development and implementation of comprehensive libraries containing fully characterized structures [14, 18, 19]. The use of libraries for structure identification relies on the reproducibility of the method and requires high repeatability and stability over time [20]. The analyses of oligosaccharides have unique issues not associated with other metabolites. Oligosaccharides have a large number of isomeric structures with very similar physical characteristics, making separation very difficult.

In this study, we present the development and optimization of a 96-well plate-based HMO extraction protocol from 107 breast milk samples collected from 45 donors, describe an automated process for nano-liquid chromatography chip time-of-flight mass spectrometry (nano-LC chip-TOF MS) peak alignment, and demonstrate rapid assignment of specific structure using a fully annotated HMO library and database that has been published previously [14, 18]. The efficiency and reproducibility of oligosaccharide extraction from milk and subsequent analysis by nano-LC chip-TOF MS are evaluated for the first time in detail. Over 250 HMO compounds are monitored, nearly 100 of which are assigned fully annotated structures, with putative structures determined for the remainder. There is no additional labeling step involved or the use of expensive commercial standards as references. Additionally, the use of time-of-flight mass spectrometry as detection allows for the identification of co-eluting peaks that can be easily distinguished based on mass. Important factors such as instrument stability, retention time reproducibility, and the linear range of the instrument response are assessed and reported herein, providing the necessary description of the validity of the presented method. Furthermore, we describe and assess the performance of an automated approach to peak alignment, allowing for the routine use of the HMO library for rapid-throughput structure identification. Together, all these improvements to our protocols drastically reduce the amount of time needed to prepare and analyze a larger number of milk samples.

Materials and methods

Human milk samples

Breast milk collected over three postpartum time points from 45 healthy mothers who gave birth to healthy term infants and enrolled in the UC Davis Foods for Health Institute Lactation Study was used in this project for method development and validation. The UC Davis Institutional Review Board approved all aspects of the study, and informed consent was obtained from all subjects. This human study was registered at clinicaltrials.gov as NCT01817127.

Materials

Reagents for Folch extraction (chloroform and methanol) and aldehyde reduction (sodium borohydride) were purchased from Sigma-Aldrich. Extract-Clean™ Carbo 150-mg bed weight porous graphitized carbon cartridges for solid-phase extraction were purchased from Grace (Deerfield, IL). For rapid-throughput format, 96-well plates were purchased from Thermo Scientific and carbon-packed plates for solid phase extraction were purchased from Glygen (Columbia, MD). Breast milk from a single donor was used to compare HMO extraction methods. Once optimized, the final rapid-throughput method was applied to a set of 107 milk samples from this cohort, with milk from 45 mothers at three time points (not all mothers donated at each time point). HMOs were extracted from a pooled milk sample from several donors and used as a standard reference. All milks were stored at −80 °C until thawed for immediate use.

Oligosaccharide extraction

A newly developed 96-well plate format for extracting unconjugated oligosaccharides from human milk was compared to our previously published method, in which oligosaccharides were extracted from each sample individually [11–13]. Both extraction methods were performed using 50-μL aliquots from the same milk sample, both done in triplicate.

As per the protocol described in our previous publications, HMOs were first extracted from milk one sample at a time. An initial volume of 50 μL of milk was diluted with 50 μL of water and defatted via centrifugation at 13,000 rcf in a cold room kept at 4 °C. The aqueous layer was recovered and subjected to an ethanol precipitation for the removal of proteins. Proteins were precipitated at −80 °C for 1 h. After 30 min of centrifugation at 13,000 rcf, 4 °C, the upper liquid fraction was collected and dried to completion. All samples were reconstituted in 50 μL of water and were reduced to alditols with 50 μL of 1 M $NaBH_4$ at 60 °C for 1.5 h, followed immediately by desalting by solid-phase extraction on graphitized carbon cartridges (SPE-GCC) with 150-mg bed weight. The desalting and SPE steps were performed on a Gilson liquid handler. Salts were removed by washing with six column volumes of deionized (DI) water (approximately 18 mL). HMOs were eluted with 6 mL of 20 % acetonitrile in water, followed by 6 mL of 40 % acetonitrile (ACN) and 0.05 % trifluoroacetic acid (TFA). SPE fractions were combined and dried by centrifugal evaporation in vacuo. Samples were then reconstituted to 50 μL with DI water and further diluted 50-fold with DI water for mass spectrometric analysis.

To achieve higher throughput, the method described above was then transferred to a 96-well plate format. Defatting, ethanol precipitation, aldehyde reduction, and solid-phase extraction were all performed on a 96-well plate. All plate centrifugation was done on a four-plate centrifuge (Eppendorf, model 5810R) with internal temperature control for 30 min at 21,130 rcf at 4 °C. All liquid handling was performed using a 12-channel pipette. Centrifugation-based solid-phase extraction was performed on 96-well plates packed with graphitized carbon. Samples were desalted with six column volumes (approximately 1.2 mL) of DI water, and HMOs were eluted with 400 μL of 20 % ACN in water and 400 μL of 40 % ACN in 0.05 % TFA. SPE fractions were combined and dried to completion in a vacuum concentrator. Each sample was then reconstituted to 50 μL with DI water, followed by a 50-fold dilution for mass spectrometry. Both extraction methods are depicted in Fig. 1.

Nano-LC chip-TOF mass spectrometry analysis

Samples from both extraction methods were analyzed on a nano-LC microfluidic chip coupled to electrospray time-of-flight mass spectrometer (nano-LC chip-TOF MS). An Agilent Technologies 1200 series HPLC, equipped with a capillary pump for sample loading, a nano-pump for separation, and an autosampler fitted for 96-well plates, was coupled to an Agilent Technologies 6210 series TOF mass spectrometer via a chip-cube interface. All samples were enriched and separated on a porous graphitized carbon chip (Agilent Technologies, model number G4240-64010), equipped with a 40-nL trapping column, and a 75-μm×43-mm analytical column. LC separation was performed with a binary solvent gradient, with aqueous solvent "A" 3 % acetonitrile/water (*v*/*v*) with 0.1 % formic acid, and organic solvent "B" 90 % acetonitrile/water (*v*/*v*) with 0.1 % formic acid. A 45-min gradient was used for the separation of glycan mixtures at a flow rate of 0.3 μL/min as follows: 0–20 min, 0–16 % B; 20–30 min, 16–44 % B; 30–35 min, 44–100 % B; 35–45 min, flush with 100 % B. The analytical column was equilibrated at the end of each gradient for 15 min with 100 % A. Time-of-flight mass spectrometry was operated in positive mode. Masses were calibrated with several internal calibrant ions over a range of masses (*m/z* 118.086, 322.048, 622.029, 922.010, 1,221.991, 1,521.972, 1,821.952, 2,121.933, 2,421.914, and 2,721.895), and a single calibrant ion (*m/z* 1,221.991) was used for reference during analysis. Data was collected using Agilent MassHunter Workstation Acquisition software, version TOF/QTOF B.02.01, and analyzed using Agilent MassHunter Qualitative Analysis, version B.03.01. Using this software, data were deconvoluted and compounds were extracted using the "Find by Molecular Features" algorithm.

Automated peak alignment and rapid, library-based structure identification

HMO peaks were rapidly identified and assigned monosaccharide compositions by matching the exact mass to a database of known HMO masses within 20 ppm (although the average mass error was typically around 5 ppm). The resulting mass lists contain neutral masses, retention times, monosaccharide compositions (denoted in this text as Hexose_GlcNAc_Fucose_Neu5Ac), and peak intensities (in ion counts). Specific oligosaccharide structures were rapidly assigned by matching mass, retention time, and elution order to an annotated HMO library with nearly 100 fully annotated structures, using the pooled milk sample as reference [14, 18]. Peak alignment was performed automatically using an in-house program written in Python 2.7 using the Scipy, Numpy, and MatPlotLib modules [21, 22]. Spectral alignment was achieved by identifying well-behaved peaks using simple criteria regarding the percent occurrence within a set and the number of closely eluting peaks, then using kernel density estimation

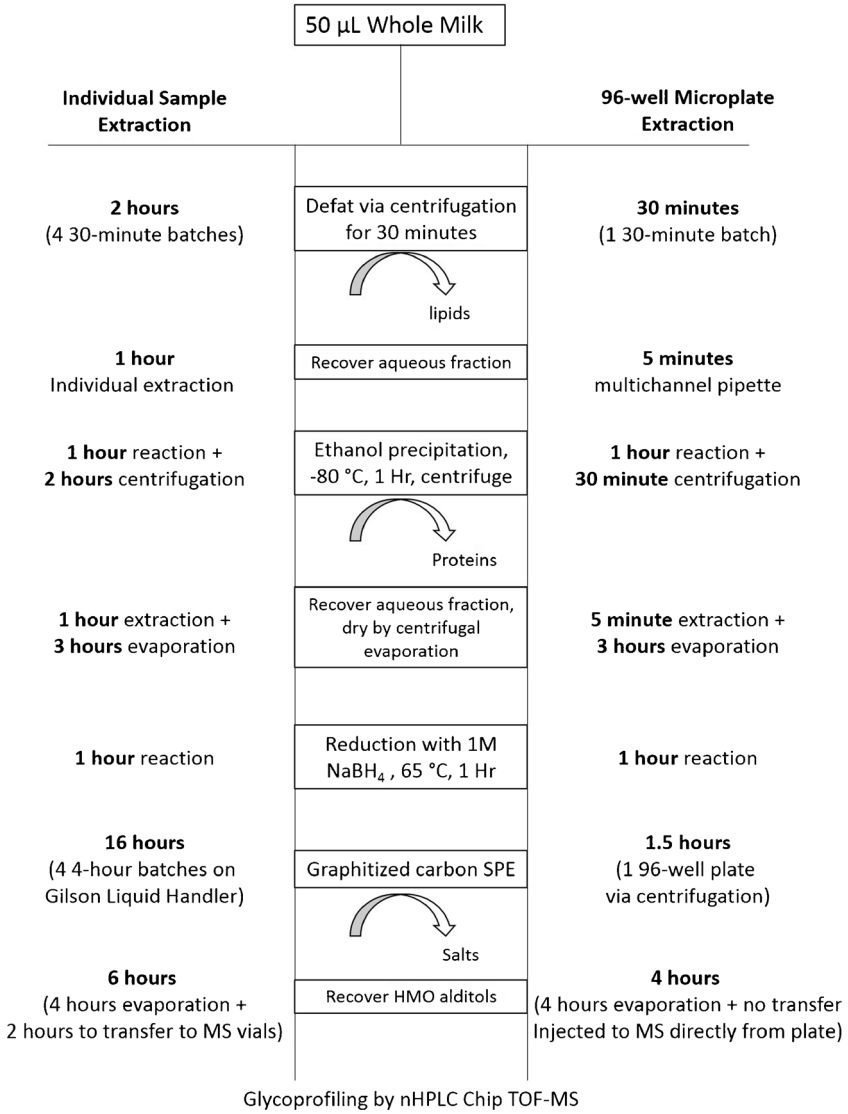

Fig. 1 Oligosaccharide extraction from milk. The *left* and *right columns* compare the previously published method by which HMOs are extracted from each sample individually (*left*) to the 96-well plate-based extraction method (*right*). Approximate timescales for processing 96 samples using each method are summarized on either side

to estimate the 95 % retention time deviation limits and set the alignment center-point for each well-behaved feature. Lastly, a linear interpolation of retention time versus retention time deviation was used to correct all retention times within each chromatogram [23]. Chromatographic peaks were aligned by iterating this process three times. Feature groups were extracted using hierarchical clustering with an estimated value of allowed retention time deviation time pairs. To compare the automated alignment program to manual data analysis, chromatographic peaks were also aligned by hand by defining a retention time range for each peak based on peak width and retention time deviation, as determined by manual inspection of the chromatograms. Each feature of the same mass within the defined retention time range was then grouped for a total peak intensity. The intensities of automatically extracted feature groups were then matched against the intensities from the manually determined groups, and percent differences in resulting peak intensities were calculated.

HMO quantitation and statistical analysis

For compositional analysis, the oligosaccharides were categorized by type: fucosylated, sialylated, and nonfucosylated neutral. In-house Python-based scripts were used to rapidly calculate the summed oligosaccharide signal (total HMO abundance) and sum the peak intensities of all fucosylated glycans, sialylated glycans, and nonfucosylated neutral glycans. The abundance of structures containing fucose *and* sialic acid were included in the sum for both the fucosylated and sialylated groups. Absolute intensities in ion counts were log10-transformed to meet homogeneity of variance assumptions. Relative abundances for each glycan type were calculated by normalizing the sum of each group to the total

oligosaccharide abundance and were expressed as percentages. To compare both extraction methods outlined in Fig. 1, means and relative standard deviations (%RSD) were calculated for the absolute and relative abundances of each glycan type. A two-tailed, unpaired student *t* test was used to compare glycan means between the two extraction methods, with an alpha value of 0.05. All statistical analyses were performed using JMP statistical software, version pro 11, and are summarized in Table 1.

Instrument reproducibility over time

To determine the reproducibility and stability of the instrument performance over time, a reference pooled milk sample was repeatedly injected over 132 h, with 12 h between each injection (approximately 5.5 days of continuous use of the instrument). Means and %RSD were calculated for total oligosaccharide abundance, the absolute and relative abundances of each glycan type, and the absolute and relative intensities of the 30 most abundant specific HMO structures.

Linear range of instrument response

The linear range of the entire 96-well plate method (from HMO extraction through LC-MS analysis) was determined by performing analysis on a serial dilution of milk. Milk was aliquoted in triplicates of decreasing volumes of milk (100, 80, 30, 60, 50, 40, 20, and 10 μL). HMOs were extracted, dried completely via centrifugal evaporation, and were reconstituted to a final volume of 100 μL with DI water, for a serial dilution ranging from one- to tenfold. The linearity of the method was tested from the initial sample preparation steps in order to assess the HMO yield of the entire method, with opportunity for variance in every step of the extraction. The total glycan signal was summed and plotted versus the starting volume of milk. The same was done with the peak intensities of chosen representative HMO structures, including lacto-N-tetraose (LNT, the most-abundant milk oligosaccharide), lacto-N-fucopentaose I (LNFP I, a secretor status marker) and sialyllacto-N-tetraose b (LST b, a lower-abundant sialylated oligosaccharide). The data sets were fitted with a least-squares regression model to determine linearity.

Results and discussion

Comparing HMO extraction methods

Prior to moving to a 96-well plate format, oligosaccharides were extracted from each sample individually, which proved to be time-consuming and required larger volumes of milk. In this study, a rapid-throughput method for HMO extraction was developed. Sample processing time and resulting LC-MS oligosaccharide profiles were compared to those of our previously published methods [11–14, 18]. An outline of the extraction steps along with a side-by-side time comparison between the old and new methods is presented in Fig. 1. Several steps in the HMO extraction procedure require centrifugation, including defatting, ethanol precipitation, and solid-phase extraction. By moving to a 96-well plate format, up to 384 samples (four 96-well plates) can be processed simultaneously. Additionally, multichannel pipettes were used for liquid handling, effectively making the liquid extraction steps 12 times faster. The most dramatic cut in sample preparation time was achieved by using centrifugation for solid-phase extraction on carbon-packed plates, making it possible to process four times as many samples in half the time. From beginning to end, HMOs can be extracted, dried down, and reconstituted for mass spectrometric analysis from 96 to 384 samples (one to four plates) in 2 days using the newly developed 96-well plate-based method, whereas it would take approximately 1 week to process the equivalent amount of samples using the previously published protocol [11–14, 18].

The glycoprofiles acquired from each method yielded only minor differences in HMO composition, and are summarized in Table 1 and graphically represented in Electronic Supplementary Material (ESM) Fig. S2. Comparable absolute amounts of total HMO, fucosylated, and sialylated

Table 1 Summarized HMO compositional analysis for the comparison of HMO extraction methods

	Method	Total HMO abundance	Fucosylation	Sialylation	Nonfucosylated neutral
Absolute intensity E+08 counts	Individual sample extraction	1.67±0.09	1.21±0.07	0.39±0.02	0.25±0.01
	96-well plate extraction	1.61±0.12	1.27±0.08	0.39±0.05	0.20±0.01
	% difference	−3.66	4.84	0.00	−22.22
Relative intensity (%)	Individual sample extraction		72.55±0.62	23.39±0.18	15.17±0.43
	96-well plate extraction	n/a[a]	79.37±0.83	24.00±1.08	12.27±0.61
	% difference		8.98	2.60	−21.16

Intensities are reported as mean ± standard deviation in E+08 counts for absolute intensity and in percentage for relative intensity

[a] Relative abundances were expressed as a percentage of the total HMO abundance

oligosaccharides were achieved by the two extraction methods, demonstrating the ability of the newly developed plate-based method to reproduce similar glycan profiles to the previously published method when starting with aliquots of the same milk. Additionally, relative percentages of fucosylation and sialylation varied by less than 10 % between the two methods. Notably, both the absolute and relative intensities of the undecorated (nonfucosylated neutral) HMO species were 20 % lower when extracted on a 96-well plate. Considering the slight increase in signal from fucosylated and sialylated HMOs in the 96-well plate method, a decrease in signal from undecorated HMOs may suggest that less of the labile bonds linking sialic acid and fucose residues were lost during sample processing. Low standard deviations in absolute and relative abundances in both methods demonstrate the ability to derive highly reproducible signal using the newly optimized plate-based extraction protocol.

Two additional steps in our previously published HMO extraction workflow were assessed for necessity in developing a rapid-throughput method: the removal of lipids by Folch liquid-liquid extraction and the precipitation of proteins in ethanol. Data from these experiment are not shown here (resulting glycoprofiles from these experiments can be found in ESM Fig. S1). However, it should be briefly noted that the Folch extraction for removing lipids from milk, although commonly used, was determined to be unnecessary for the oligosaccharide extraction process and was not used in the development and optimization of the plate-based method. It was also determined that the ethanol precipitation of proteins could be omitted without altering the HMO profiles; however, it was discovered that when this step was removed, there was clogging of the microfluidic column by residual impurities after approximately 100 injections, making it difficult to maintain a stable current in the electrospray ionization process during the analysis of large-scale sample sets. For these reasons, this step was retained in the HMO extraction protocol. Additionally, the use of graphitized carbon for desalting prior to mass spectrometric analysis could possible contribute to the loss of smaller, di- and tri-saccharides when analyzing complicated glycan mixtures, such as lactose and fucosyllactose.

The resulting glycoprofile from the reference pooled milk sample using the optimized 96-well plate extraction is shown in Fig. 2. The overlaid extracted compound chromatogram demonstrates the extensive and complex glycome of a milk sample using the plate-based method. Typically 100–150 HMO species are monitored in a single milk sample and up to 300 in pooled milk. Specific structures were assigned based on mass and retention time using the annotated HMO library, and putative compositions (shown in black) were assigned based on accurate mass. A list of monitored HMO species and identified specific structures from the reference HMO pool can be found in ESM Table S1. A list of full HMO names is provided in ESM Table S2. See previously published HMO library references for full structures [14, 18].

Instrument stability and reproducibility over time

When analyzing large sample sets using nano-LC chip-TOF MS, it is important that the instrument is stable over time. To evaluate method reproducibility, a reference HMO pool was analyzed intermittently between samples every 12 h over the course of 132 h—approximately 5.5 days of continuous use. The peak intensity and retention time %RSDs are tabulated in Table 2. The overlaid extracted ion chromatogram in Fig. 3 demonstrates the reproducibility of several abundant peaks, including the neutral (nonsialylated) isomers of *m/z* 710.27, *m/z* 856.3, *m/z* 684.25, *m/z* 491.20, 538.20, and *m/z* 611.23 (Fig. 3a), and several anionic (sialylated) species, including the isomers of *m/z* 683.74, *m/z* 1,001.35, *m/z* 1,001.35, and *m/z* 636.23 (Fig. 3b).

Retention times were highly reproducible for both neutral and anionic species, with all RSDs <0.5 %, allowing for structure assignment based on retention time matches to the library. Sialylated species generally showed more shifting in retention time than neutral species, but still demonstrated excellent reproducibility within an RSD <0.5 %. The reproducibility of absolute peak intensity (log10-transformed) was also precise, with RSD ranging from 0.4 % to 6.9 %, with the majority RSD ≤3 %. Relative abundances normalized to the total HMO signal were generally reproducible, yielding RSD ranging from 2.1 to 24.6 %, with the majority being under 15 %. The intensity of *m/z* 1,220.45 peak at 13.77 min demonstrated the largest deviation with a RSD=65 %, which was an anomalous occurrence, as it was the only peak out of 30 that varied drastically. Generally, the less-abundant oligosaccharides demonstrated greater variability in peak intensity, but still yielded RSD ≤20 %. The overall performance of the nano-LC chip-TOF MS over the course of 5.5 days of continuous use proved to be highly reproducible, as was also observed in a previous study by Ruhaak et al. [24], which report high reproducibility in peak areas for inter- and intraday variations (CV=3.9 and 2.1 %, respectively) in N-glycan analysis using the same nano-LC-MS instrumentation. These results show that the retention times and peak intensities are stable enough over time to use with an HMO library as a means of structure identification and quantitation in large sample sets.

Linearity of HMO glycoprofiling method

The linearity of the 96-well plate method was tested from the initial sample preparation steps in order to assess the HMO yield of the entire method, with opportunity for variance in every step of the extraction. The total glycan signal was summed and plotted versus the starting volume of milk, as

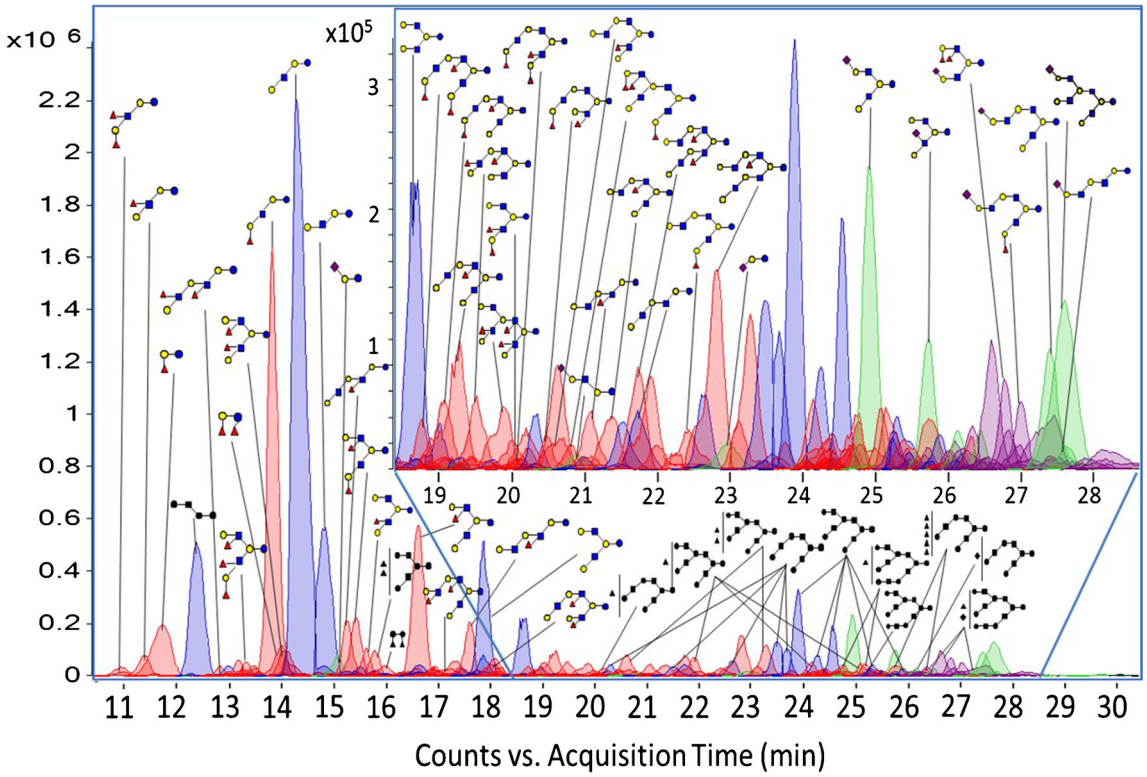

Fig. 2 Overlaid extracted compound chromatogram acquired from nano-LC chip-TOF MS analysis. Each glycan class is designated with a different color: nonfucosylated neutral (*red*), fucosylated (blue), sialylated (*green*),, and fucosylated (*purple*). The *Y*-axis represents peak intensity reported in ion counts as determined by time-of-flight mass spectrometry. The *X*-axis reports retention time in minutes

shown in Fig. 4a. The same was done with the peak intensities of chosen representative HMO structures, including LNT (the most-abundant milk oligosaccharide), LNFP I (a fucosylated HMO used as a marker for secretor status), and LST b (a lower-abundant sialylated oligosaccharide), plotted in Fig. 4b–d. The data sets were fitted with a least-squares regression model to determine linearity, with the linear equation and coefficient of determination (R^2) value inset. All plots in Fig. 4 suggest that the method is linear from 10- to 50-μL starting volume of milk. The concentration of HMO can be estimated from the starting volume of milk by assuming that the average HMO concentration is 10 mg/mL, which is a typical value reported in literature for mature (>2 weeks) milk [1]. It can be estimated that beginning with 10–100 μL of milk, reconstituting the dry isolated oligosaccharides to 100 μL and diluting 50-fold for LC-MS analysis yields a range of approximately 0.02–0.2 mg/mL of total HMO concentration. Figure 4a plots instrument response versus starting milk volume and demonstrates that the complete method, including sample preparation, is linear (R^2=0.942) between 10-μL (approximately 0.02 mg/mL) and 50-μL (approximately 0.1 mg/mL) starting volume of milk. A large portion of the total oligosaccharide concentration can be attributed to the large amounts of LNT. Due to the large range of concentrations among the hundreds of HMO species, an optimal concentration must be used in order to detect the very low-abundant species without saturating the PGC columns with LNT, which usually means working with concentrations well above the limit of detection. In this study, signals from the low-abundant HMOs were still detected at the lowest total HMO concentration when starting with 10 μL of milk (0.1 μg from a 1-μL injection). The limiting step in this method is solid-phase extraction on carbon 96-well plates, which reaches a saturation point when more than 50 μL of milk is used (approximately 0.5 mg of total HMO). As all the plots in Fig. 4 suggest, HMO concentrations in 60, 80, and 100 μL of milk are not within the linear range of the method and demonstrate a loss of signal that is indicative of saturating the SPE columns.

Peak alignment

HMO peaks were rapidly identified by filtering the LC-MS output against a library of exact masses and monosaccharide compositions. The HMO reference pool sample contained 45 unique compositions, with 200 to 250 unique structures. Fully annotated structures are then assigned based on retention time, making retention time reproducibility essential for rapid HMO identification in large sample sets. Although retention times have been shown to be reproducible in repeated injects of the reference pool, more shifting is found when comparing non-identical injections of different samples. An in-house, python-based program has been developed for rapid peak alignment

Table 2 Summary of nano-LC chip-TOF MS stability and reproducibility in repeated injections of the reference HMO pooled sample

Database mass	Average observed mass (Da)	Average mass error (ppm)	Monosaccharide composition[a]	Average retention time (min)	Retention time %RSD	Average intensity (counts)	Average intensity log10	Intensity %RSD	Relative average intensity (% of total HMO)	Relative intensity %RSD
490.190	490.19	2.59	2_0_1_0	9.80	0.18	7.35E+06	6.87	0.40	4.32	14.05
635.227	635.23	1.54	2_0_0_1	19.88	0.17	3.60E+05	5.55	0.84	0.19	7.20
636.248	636.25	0.94	2_0_2_0	12.17	0.19	2.53E+06	6.40	0.91	1.47	3.47
709.264	709.26	0.81	3_1_0_0	12.49	0.16	5.51E+07	7.74	0.68	31.94	2.10
855.322	855.32	0.47	3_1_1_0	8.90	0.09	1.55E+06	6.19	0.74	0.92	20.44
855.322	855.32	0.81	3_1_1_0	9.31	0.12	4.05E+06	6.61	0.41	2.37	10.83
855.322	855.32	1.03	3_1_1_0	12.05	0.13	1.99E+07	7.30	0.56	11.61	6.04
1,000.359	1,000.36	3.84	3_1_0_1	18.50	0.31	2.91E+05	5.46	1.66	0.17	9.08
1,001.380	1,001.38	0.98	3_1_2_0	8.73	0.13	1.58E+06	6.20	0.63	0.93	13.62
1,074.396	1,074.40	1.80	4_2_0_0	16.12	0.24	5.79E+06	6.75	1.35	3.33	8.73
1,074.396	1,074.40	2.21	4_2_0_0	16.87	0.26	2.50E+06	6.39	1.41	1.44	9.63
1,074.396	1,074.40	−0.03	4_2_0_0	20.12	0.24	7.75E+05	5.88	1.88	0.44	12.58
1,220.454	1,220.47	9.45	4_2_1_0	13.43	0.18	2.54E+06	6.40	1.36	1.46	11.60
1,220.454	1,220.46	1.23	4_2_1_0	13.77	0.16	5.21E+05	5.60	6.94	0.27	65.17
1,220.454	1,220.47	12.50	4_2_1_0	14.80	0.26	7.45E+06	6.87	1.01	4.30	5.46
1,220.454	1,220.46	5.51	4_2_1_0	15.80	0.26	1.62E+06	6.20	1.34	0.93	8.91
1,220.454	1,220.46	2.27	4_2_2_0	19.03	0.22	8.18E+05	5.90	2.20	0.47	19.08
1,365.492	1,365.49	−0.53	4_2_0_1	23.08	0.26	7.13E+05	5.84	2.20	0.41	16.42
1,365.492	1,365.49	0.18	4_2_0_1	23.96	0.21	2.55E+05	5.39	2.65	0.14	18.50
1,366.512	1,366.51	1.35	4_2_2_0	10.90	0.14	3.89E+05	5.58	1.38	0.23	12.51
1,366.512	1,366.51	1.39	4_2_2_0	11.43	0.15	6.33E+05	5.80	1.03	0.37	10.76
1,366.512	1,366.51	1.40	4_2_2_0	12.20	0.17	1.67E+06	6.22	1.17	0.97	7.04
1,366.512	1,366.51	1.58	4_2_2_0	13.69	0.17	3.42E+06	6.52	1.54	1.96	12.20
1,366.512	1,366.52	3.38	4_2_2_0	15.16	0.21	2.05E+05	5.30	2.22	0.12	15.70
1,366.512	1,366.51	0.71	4_2_2_0	18.14	0.34	3.80E+05	5.57	1.68	0.22	9.47
1,439.528	1,439.53	0.54	5_3_0_0	22.06	0.26	1.42E+06	6.14	1.79	0.81	13.89
1,511.550	1,511.55	0.69	4_2_1_1	21.50	0.41	6.83E+04	4.82	2.52	0.04	24.63
1,511.550	1,511.55	0.07	4_2_1_1	23.45	0.26	8.53E+04	4.92	1.73	0.05	15.62
1,512.570	1,512.57	0.51	4_2_3_0	11.26	0.16	7.48E+05	5.87	1.18	0.44	12.53
1,512.570	1,512.57	1.37	4_2_3_0	12.19	0.20	1.06E+06	6.02	1.23	0.61	11.32

[a] Monosaccharide compositions are denoted as Hexose_GlcNac_Fucose_Neu5Ac

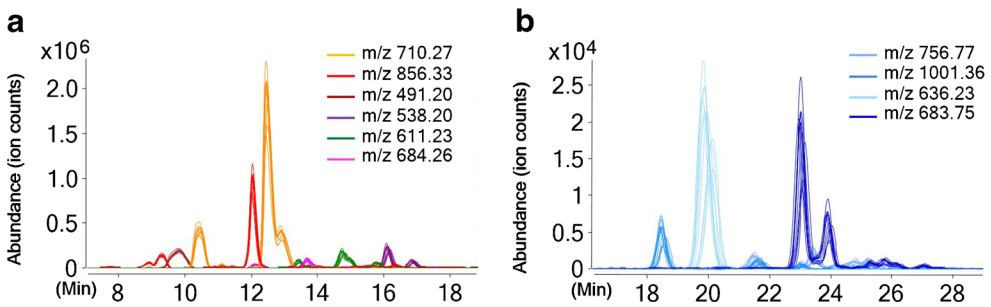

Fig. 3 Stability and reproducibility of nano-LC chip-TOF mass spectrometry method was demonstrated with repeated injections of a reference milk pool every 12 h over the course of 153 h. Repeatability of retention time and abundance are demonstrated with overlaid extracted ion chromatograms for the most abundant neutral (**a**) and anionic (**b**) oligosaccharide species. The *Y*-axis represents peak intensity reported in ion counts as determined by time-of-flight mass spectrometry. The *X*-axis reports retention time in minutes

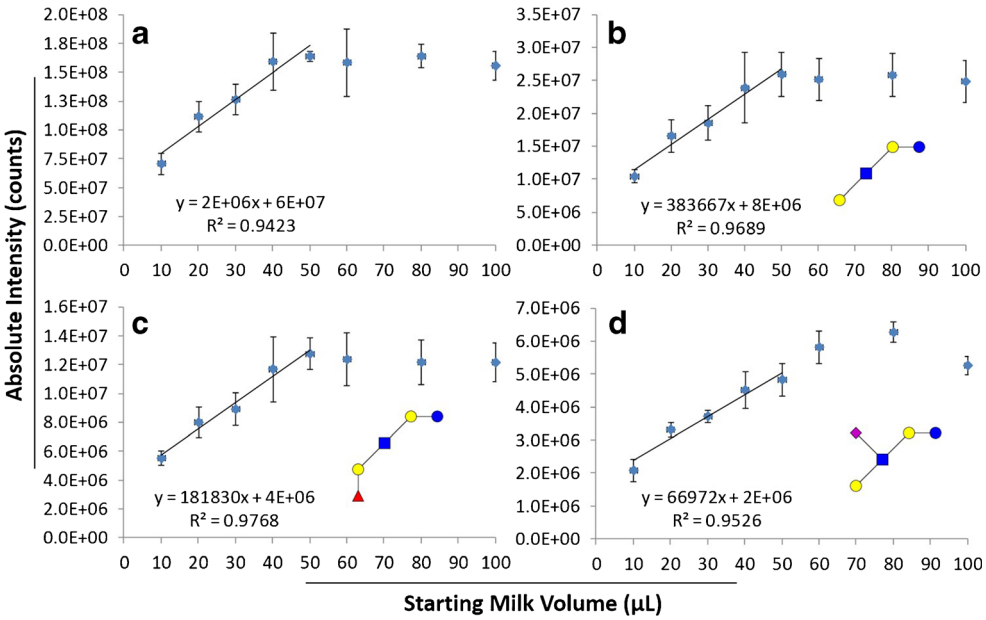

Fig. 4 Linearity of instrument response as a function of oligosaccharide concentration. Total HMO intensity (**a**), lacto-N-tetraose (LNT) (**b**), lacto-N-fucopentaose (LNFP I) (**c**), and sialyllacto-N-tetraose (LST b) (**d**) yield a linear instrument response when up to 50 μL of milk is used for analysis. *Error bars* are represented by standard deviations. In all four plots, the *Y*-axis represents peak intensity reported in ion counts as determined by time-of-flight mass spectrometry. The *X*-axis reports the initial starting volume of whole milk (before sample preparation) in microliters

to automatically correct for retention time deviations. Prior to the development of this in-house software, peak alignment and retention time correction were done manually, as described in previous publications on HMOs from our group [11, 12, 14, 18, 25, 26]. Manually compiling a list of identified features and their abundances was a particularly time-consuming step, having to address issues with retention time shifting of closely eluting isomers and chromatographic peak splitting during integration (multiple features representing the same compound) for each individual sample. Data analysis done in this manner is not ideal for large samples sets due to the immense amount of time it would take to manually align hundreds of HMO peaks in hundreds of samples.

To demonstrate the comparability of the manual and automated methods for peak alignment, HMOs were extracted using the plate-based method described above from the 107 breast milk samples used in this project and the resulting deconvoluted data were aligned both manually and automatically. The differences in peak intensities resulting from manual and automated analyses were determined for two sets of closely eluting isomers—one set of neutral glycans, *m/z* 1 221.45, monosaccharide composition 4_2_1_0 (denoted as Hexose_GlcNac_Fucose_Neu5Ac) shown in Fig. 5a, and the other set of acidic glycans with *m/z* 1,366.75, monosaccharide composition 4_2_0_1 shown in Fig. 5c.

The isomers of *m/z* 1,221.45 were chosen to demonstrate the ability of the alignment program to discriminate between isomeric peaks that elute within 0.5 to 1.0 min of each other and to correctly group their abundances as part of the same oligosaccharide structure. Six major isomers were observed for *m/z* 1,221.45. These peaks were identified according to the HMO library as MFpLNH IV, 4210a, MFLNH III, MFLNH I, IFLNHIII, and IFLNH I, in order of retention time and are labeled in Fig. 5a as peaks a–e, respectively (peak "c" represents the co-eluting isomers MFLNH III/I). The percent difference in manual versus automated peak intensities was determined for each isomer in 107 samples and plotted in Fig. 5b. Approximately 90 % of the points for each isomer fall into the ±10 % difference zone, as highlighted by the gray band. In fact, between 60 and 80 % of the points for each isomer demonstrated 0 % difference between the peak intensities aligned and grouped manually compared to those aligned and grouped automatically (59 % in 1221a, 81 % in 1221b, 82 % in 1221c, 58 % in 1221d, and 81 % in 122e). A positive percent difference indicates a peak intensity determined to be greater in the automatically aligned data than in the manual data. Upon inspection of these deviations, it was determined that in a small portion of the points for each isomer (≤9 %), the automated alignment program would either incorrectly combine two features (creating a negative percent difference for another isomer) or would correctly incorporate features that were overlooked or otherwise calculated incorrectly by hand. The same trend was observed when examining the agreement between automatically and manually determined peak intensities of the sialylated isomers of *m/z* 1,366.75. As can be seen in Fig. 5c, four major isomers were observed for *m/z* 1,366.75, identified by the library as S-LNH, 4021a, S-LNnH II, and 4021b, in order of retention time and labeled a–d, respectively. This particular set of isomers was chosen for assessing the abilities of the automated alignment program due to the three isomers that elute within less than 0.5 min of each other and the visible shift in retention time. Again, over 90 % of the points for each isomer were within ±10 % difference, with the majority yielding identical abundances (0 % difference). This provides sufficient evidence that the automated alignment program can correctly follow a series

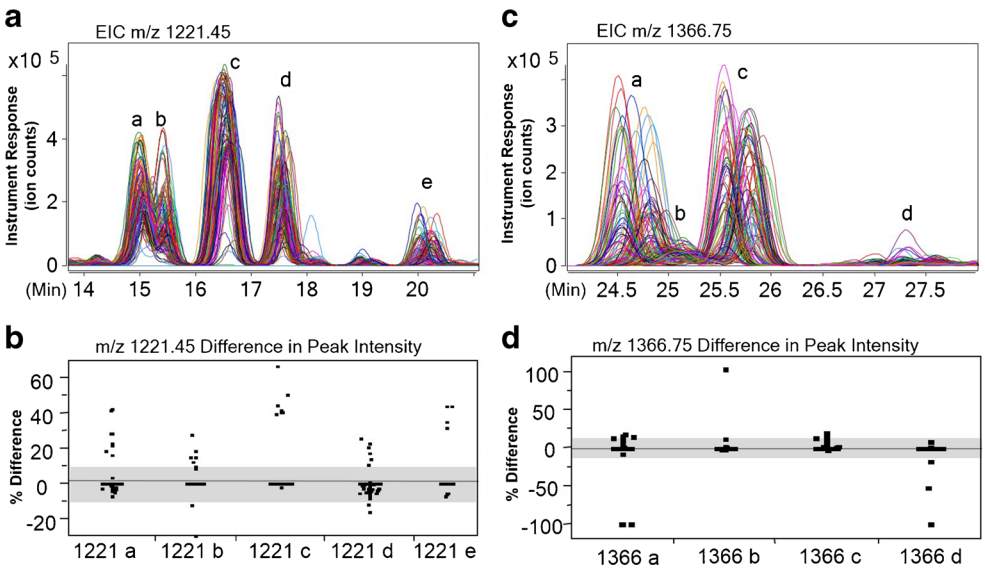

Fig. 5 Overlaid extracted ion chromatograms of *m/z* 1,221.45 (**a**) and *m/z* 1,366.75 (**c**) demonstrate closely eluting isomers and the extent of retention time shifting in sialylated species. Each isomer of a given *m/z* was lettered in order of elution. Scatterplots were used to visualize differences in peak intensities between manual and automatic peak alignment and feature grouping for both neutral *m/z* 1,221.45 isomers (**b**) and anionic *m/z* 1,366.75 isomers (**d**). The *gray bands* highlight the ±10 % range, where the majority of the points fall for each isomer. The *X*-axis on plots **b** and **d** are HMO isomers labeled with the *m/z* and the corresponding letter from their chromatographic elution order shown in plots **a** and **c**

of decision-making processes to correctly group features together to determine an accurate peak intensity in a diverse mixture of complex glycans and can do so with efficacy and in excellent agreement with manual data analysis (Fig. 5d).

Conclusion

In this study, a concept for rapid-throughput analysis of oligosaccharides is presented using robust and reproducible nano-LC chip-TOF MS method for isomer-specific milk glycan analysis. The method includes the initial sample preparation steps all the way through automated data analysis. The successful transition to a 96-well plate format for the extraction of unconjugated oligosaccharides from milk was optimized for sensitivity and time efficiency. Superfluous steps were removed (e.g., Folch liquid-liquid extraction for the removal of lipids), and remaining steps were optimized for use on a 96-well plate without compromising the glycoprofiles. By moving to a 96-well plate format, a considerable amount of time is cut from the sample preparation process, allowing for hundreds of samples to be processed with rapid throughput. Although not fully automated, we significantly improve upon previous HMO isolation protocols, by incorporating the use of plate centrifugation to perform solid-phase extraction to up to four 96-well plates simultaneously. Our approach to glycan analysis is also unique in the application of a fully annotated HMO library for profiling and structure identification, demonstrating the ability for rapid- and potentially high-throughput glycomic studies. Recent efforts by Kottler et al. have initiated the use of an HMO library for application in their multiplexed capillary gel electrophoresis techniques, demonstrating an effective use of HMO "fingerprints" for structure identification [16]. In our study, by matching retention time and accurate mass to a database consisting of nearly 100 previously elucidated milk oligosaccharide structures, a more comprehensive view of the glycome is achieved without having to rely on the few HMO standards that are commercially available. Although the use of standards provides more accurate quantitation, as has been demonstrated in a number of studies, including a statistically rigorous study from Bao et al. [15], the chip-based nano-LC-TOF MS method described here yields an equally sensitive method that yields linear and reproducible data for extensive glycan analysis. Additionally, the stability of the instrumentation's performance over long periods of continuous use is ideal for large-scale, rapid-throughput studies. Furthermore, with the use our newly developed python-based scripts, peak alignment and retention time correction are now automated, accurate, and reproducible, allowing for the instantaneous alignment of hundreds of samples.

Acknowledgments The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Research reported in this publication was supported by the National Institute of Child Health and Human Development, National Institute of General Medicine, and National Center of Complementary and Alternative Medicine of the National Institutes of Health under award numbers R01HD061923, R01GM049077, R01 AT007079, and 1U24DK097154.

References

1. Bode L (2012) Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology 22(9):1147–1162. doi:10.1093/glycob/cws074
2. Bode L, Kuhn L, Kim H-Y, Hsiao L, Nissan C, Sinkala M, Kankasa C, Mwiya M, Thea DM, Aldrovandi GM (2012) Human milk oligosaccharide concentration and risk of postnatal transmission of HIV through breastfeeding. Am J Clin Nutr 96(4):831–839. doi:10.3945/ajcn.112.039503
3. Kunz C, Rudloff S, Baier W, Klein N, Strobel S (2000) Oligosaccharides in human milk: structural, functional, and metabolic aspects. Annu Rev Nutr 20(1):699–722. doi:10.1146/annurev.nutr.20.1.699
4. Morrow AL, Ruiz-Palacios GM, Jiang X, Newburg DS (2005) Human-milk glycans that inhibit pathogen binding protect breast-feeding infants against infectious diarrhea. J Nutr 135(5):1304–1307
5. Newburg DS, Ruiz-Palacios GM, Morrow AL (2005) Human milk glycans protect infants against enteric pathogens. Annu Rev Nutr 25(1):37–58. doi:10.1146/annurev.nutr.25.050304.092553
6. Hua S, Lebrilla C, An HJ (2011) Application of nano-LC-based glycomics towards biomarker discovery. Bioanalysis 3(22):2573–2585. doi:10.4155/bio.11.263
7. Costello CE, Contado-Miller JM, Cipollo JF (2007) A glycomics platform for the analysis of permethylated oligosaccharide alditols. J Am Soc Mass Spectrom 18(10):1799–1812. doi:10.1016/j.jasms.2007.07.016
8. Erney RM, Malone WT, Skelding MB, Marcon AA, Kleman-Leyer KM, O'Ryan ML, Ruiz-Palacios G, Hilty MD, Pickering LK, Prieto PA (2000) Variability of human milk neutral oligosaccharides in a diverse population. J Pediatr Gastroenterol Nutr 30(2):181–192
9. Leo F, Asakuma S, Nakamura T, Fukuda K, Senda A, Urashima T (2009) Improved determination of milk oligosaccharides using a single derivatization with anthranilic acid and separation by reversed-phase high-performance liquid chromatography. J Chromatogr A 1216(9):1520–1523. doi:10.1016/j.chroma.2009.01.015
10. Thurl S, Henker J, Siegel M, Tovar K, Sawatzki G (1997) Detection of four human milk groups with respect to Lewis blood group dependent oligosaccharides. Glycoconj J 14(7):795–799. doi:10.1023/A:1018529703106
11. De Leoz M, Wu S, Strum J, Niñonuevo M, Gaerlan S, Mirmiran M, German JB, Mills D, Lebrilla C, Underwood M (2013) A quantitative and comprehensive method to analyze human milk oligosaccharide structures in the urine and feces of infants. Anal Bioanal Chem 405(12):4089–4105. doi:10.1007/s00216-013-6817-1
12. Ninonuevo MR, Park Y, Yin H, Zhang J, Ward RE, Clowers BH, German JB, Freeman SL, Killeen K, Grimm R, Lebrilla CB (2006) A strategy for annotating the human milk glycome. J Agric Food Chem 54(20):7471–7480. doi:10.1021/jf0615810
13. Strum JS, Kim J, Wu S, De Leoz MLA, Peacock K, Grimm R, German JB, Mills DA, Lebrilla CB (2012) Identification and accurate quantitation of biological oligosaccharide mixtures. Anal Chem 84(18):7793–7801. doi:10.1021/ac301128s
14. Wu S, Grimm R, German JB, Lebrilla CB (2010) Annotation and structural analysis of sialylated human milk oligosaccharides. J Proteome Res 10(2):856–868. doi:10.1021/pr101006u
15. Bao Y, Chen C, Newburg DS (2013) Quantification of neutral human milk oligosaccharides by graphitic carbon high-performance liquid chromatography with tandem mass spectrometry. Anal Biochem 433(1):28–35. doi:10.1016/j.ab.2012.10.003
16. Kottler R, Mank M, Hennig R, Müller-Werner B, Stahl B, Reichl U, Rapp E (2013) Development of a high-throughput glycoanalysis method for the characterization of oligosaccharides in human milk utilizing multiplexed capillary gel electrophoresis with laser-induced fluorescence detection. Electrophoresis 34(16):2323–2336. doi:10.1002/elps.201300016
17. Blank D, Gebhardt S, Maass K, Lochnit G, Dotz V, Blank J, Geyer R, Kunz C (2011) High-throughput mass finger printing and Lewis blood group assignment of human milk oligosaccharides. Anal Bioanal Chem 401(8):2495–2510. doi:10.1007/s00216-011-5349-9
18. Wu S, Tao N, German JB, Grimm R, Lebrilla CB (2010) Development of an annotated library of neutral human milk oligosaccharides. J Proteome Res 9(8):4138–4151. doi:10.1021/pr100362f
19. Aldredge D, An HJ, Tang N, Waddell K, Lebrilla CB (2012) Annotation of a serum N-glycan library for rapid identification of structures. J Proteome Res 11(3):1958–1968. doi:10.1021/pr2011439
20. Hua S, An HJ, Ozcan S, Ro GS, Soares S, DeVere-White R, Lebrilla CB (2011) Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers. Analyst 136(18):3663–3671. doi:10.1039/C1AN15093F
21. Jones E, Oliphant T, Peterson P (2001) Scipy: open source scientific tools for Python. http://www.scipy.org. Accessed 07 Jan 2014
22. Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9(3):90–95
23. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78(3):779–787. doi:10.1021/ac051437y
24. Ruhaak LR, Taylor S, Miyamoto S, Kelly K, Leiserowitz G, Gandara D, Lebrilla C, Kim K (2013) Chip-based nLC-TOF-MS is a highly stable technology for large-scale high-throughput analyses. Anal Bioanal Chem 405(14):4953–4958. doi:10.1007/s00216-013-6908-z
25. Tao N, Wu S, Kim J, An HJ, Hinde K, Power ML, Gagneux P, German JB, Lebrilla CB (2011) Evolutionary glycomics: characterization of milk oligosaccharides in primates. J Proteome Res 10(4):1548–1557. doi:10.1021/pr1009367
26. Totten SM, Zivkovic AM, Wu S, Ngyuen U, Freeman SL, Ruhaak LR, Darboe MK, German JB, Prentice AM, Lebrilla CB (2012) Comprehensive profiles of human milk oligosaccharides yield highly sensitive and specific markers for determining secretor status in lactating mothers. J Proteome Res 11(12):6124–6133. doi:10.1021/pr300769g