Published in final edited form as:

Anal Bioanal Chem. 2019 September ; 411(23): 5937–5949. doi:10.1007/s00216-019-01997-7.

Accurate mass and retention time library of serum lipids for type 1 diabetes research

Ngoc Vu^{1,2}, Monica Narvaez-Rivas², Guan-Yuan Chen², Marian J. Rewers³, Qibin Zhang^{1,2}

¹Department of Chemistry & Biochemistry, University of North Carolina at Greensboro, Greensboro, NC 27412, USA

²Center for Translational Biomedical Research, University of North Carolina at Greensboro, North Carolina Research Campus, Kannapolis, NC 28082, USA

³Barbara Davis Center for Diabetes, University of Colorado School of Medicine, Aurora, CO 80045, USA

Abstract

Dysregulated lipid species are linked to various disease pathologies and implicated as potential biomarkers for type 1 diabetes (T1D). However, it is challenging to comprehensively profile the blood specimen lipidome with full structural details of every lipid molecule. The commonly used reversed-phase liquid chromatography-tandem mass spectrometry (RPLC-MS/MS)-based lipidomics approach is powerful for the separation of individual lipid species, but lipids belonging to different classes may still co-elute and result in ion suppression and misidentification of lipids. Using offline mixed-mode and RPLC-based two-dimensional separations coupled with MS/MS, a comprehensive lipidomic profiling was performed on human sera pooled from healthy and T1D subjects. The elution order of lipid molecular species on RPLC showed good correlations to the total number of carbons in fatty acyl chains and total number of double bonds. This observation together with fatty acyl methyl ester analysis was used to enhance the confidence of identified lipid species. The final T1D serum lipid library database contains 753 lipid molecular species with accurate mass and RPLC retention time uniquely annotated for each of the species. This comprehensive human serum lipid library can serve as a database for high-throughput RPLC-MS-based lipidomic analysis of blood samples related to T1D and other childhood diseases.

Keywords

Human serum lipidome; Type 1 diabetes; Lipid profiling; Mixed-mode LC; RPLC-MS/MS; Accurate mass and time tag

Qibin Zhang, q_zhang2@uncg.edu.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00216-019-01997-7>) contains supplementary material, which is available to authorized users.

Compliance with ethical standards This research analyzed de-identified human serum samples collected from clinical studies and has IRB approval as stated in the “Materials and methods” section. All authors have read and agreed to the final version of this manuscript.

Conflict of interest The authors declare that they have no competing interests.

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

Introduction

Besides of being essential structural components of cell membrane, lipids have other distinctive biochemical roles in providing a hydrophobic environment for membrane proteins, assisting cell signaling process, regulating action of hormones, and storing biochemical energy [1, 2]. Dysregulation of lipid metabolism has offered critical insights into the pathogenesis of complex diseases, and lipids are identified as biomarkers to cancers, diabetes, and Alzheimer's and other inflammatory diseases [3, 4]. While clinical type 1 diabetes (T1D) features metabolic dysregulation of some serum lipid species, interestingly, changes in lipidome appear to precede hyperglycemia or even the appearance of islet autoimmunity [5–7].

Biomarkers can be either secreted or leaked from pathologic tissues to bloodstream. Although cell, tissue, and biofluid samples are routinely used to study diseases, blood plasma and serum are the most commonly used specimen for clinical diagnostics because of their availability [8]. However, significant analytical challenges are associated with this type of complex specimen, namely the identification of the low abundant lipids and the differentiation of isobaric and isomeric species commonly exist in glycerophospholipids and glycerolipids classes. Because of these, most of the lipids reported in the literature were characterized to the level of summed composition, i.e., total number of carbon and double bonds in the fatty acyls, which results in ambiguity in characterizing the exact molecular structure and, in turn, hinders further investigation into the roles of these lipids in biological processes [9].

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is increasing used for untargeted lipidomic analysis owing to the increases in LC resolving power, decreasing particle size, novel column chemistry, and new separation mechanisms [10–15]. The popularity of RPLC-MS-based method can be further explained by several advantages such as more reliable identifications of individual lipid species exist at trace level and the separation of isomeric and isobaric lipids with reduced ion suppression [11, 12, 16]. Software, such as LipidBlast, LipidSearch, LipidMiner, and MzMine2, were developed to process LC-MS/MS-based untargeted lipidomics data and to handle the need of automated data processing [17–20]. These software use in silico generated mass fragment libraries—many based on LIPID MAPS (<https://www.lipidmaps.org/>)—for lipid spectrum annotation. The drawback of providing a large database for lipidomic profiling in compound identification is the prevalence of false-positive identifications. Incorporating LC retention time information into the database search could greatly reduce the search space and improve the search accuracy. In this respect, accurate mass and time tag approach has been developed for proteomics [21, 22] and more recently applied to metabolomics [23], but rarely this approach has been used in lipidomics [24].

Previously, we developed an offline two-dimensional LC-MS/MS method for untargeted lipidomic profiling [12]. In this approach, a mixed-mode LC and RPLC coupled to a high-resolution mass spectrometer was demonstrated to double the lipidomic coverage for complex tissue and plasma samples in comparison to RPLCMS/MS alone. More importantly, very reproducible retention time and high mass measurement accuracy was

achieved for each lipid molecular species. In current work, we apply this approach for a comprehensive identification of lipids in T1D patient sera. Total lipids extracted from pooled sera were fractionated using mixed-mode LC based on the head group of each lipid class; collected fractions were further separated on a RPLC-MS/MS platform established in our laboratory, composed of C30 column and a high-resolution, fast scanning bench-top Q-Orbitrap mass spectrometer. In total, 753 lipid molecular species were confidently identified and used to build an in-house lipid library with each lipid species annotated with LC retention time and accurate mass.

Materials and methods

Human serum samples and total lipid extraction

De-identified serial human serum samples were obtained from the Diabetes Auto Immunity Study in the Young (DAISY) cohort. The DAISY study protocol was approved by the Institutional Review Board of the University of Colorado and detailed study design and methods have been previously published [25, 26]. Written informed consent were obtained for all study participants from a parent or legal guardian. Samples were stored at -80°C prior to analysis. Analysis of these samples was also approved by the Institutional Review Board of the University of North Carolina at Greensboro. All research was performed in accordance with relevant guidelines/regulations.

In total, serum samples from 50 subjects were selected from two groups: type 1 diabetes (T1D) group comprised of children who developed islet autoimmunity and progressed to T1D and the control group from children who remained negative islet autoimmunity at all times. Four time points (0.7 to 14.7 years of age) during the disease progression in the T1D group were selected for this study, which include the earliest time point possible in the longitudinal study, the time point prior to and after appearance of persistent islet autoimmunity, and the time point of clinical diagnosis. Sample selection of the control group was age and sex matched to the T1D group.

Pooled samples were created to represent different stages of disease development and also reflect the nature of each group. Aliquot of 5 μL from each sample within the first two time points of each group were pooled, so were the last two time points. Extraction of serum lipids was carried out following a modified Folch method [12]. Briefly, pooled samples were diluted with cold (-20°C) chloroform/methanol (2:1, v/v) at ratio of 5:1 (solvent/sample ratio). The mixture was vortexed for 10 s, then set at room temperature for 10 min and vortexed again before centrifuging for 10 min at rate 10,000 RPM. The chloroform phase was collected to a glass vial, followed by evaporation of the extracted lipids to dryness under vacuum, and stored at -80°C in nitrogen gas prior to further analysis.

Chemical reagents and standards

Ammonium formate, *n*-heptane, acetone, methanol, isopropanol (IPA), and water (Optima® LC-MS grade) were provided by Fisher Scientific (Fair Lawn, NJ). Acetonitrile (ACN) and formic acid were of LC/MS quality and acquired from Fluka (Germany). Chloroform (HPLC grade) and ammonia solution were purchased from Merck (Germany).

Lipid standards were purchased from Avanti Polar Lipids (Alabaster, AL), which include Cer d18:0/24:1, PC 18:0/18:1, PE 16:0/18:1, PG 16:0/18:1, PS 16:0/18:1, PI 21:0/22:6, PC P-18:0/20:4, PC O-18:0/20:0, PE P-18:0/22:6, PE O-18:0/18:0, SM d18:1/12:0, and d5-TG ISTD Mix I. The acylglycerols, fatty acids, cholesterol, and cholesterol esters were obtained from Nu-Chek Prep, Inc. (Elysian, MN).

Mixed-mode LC fractionation of lipid classes

The mixed-mode LC separation was performed according to our previous published method using an Agilent HPLC equipped with a quaternary pump and an Agilent 1260 Infinity evaporative light scattering detector (ELSD) (Palo Alto, CA, USA) [12]. The method was run on a Chromolith Performance Si column (100 mm × 4.6 mm, macropore size 2.1 μm and mesopore size 13 nm, Merck, Darmstadt, Germany). The autosampler was set up at 23 °C and the injection volume was 10 μL, equivalent to 100 μL of serum.

Overall, all lipid classes were collected into these fractions with the following order: CE and TG from 1 to 3.2 min; Chol and 1,3-DG from 3.21 to 4.1; 1,2-DG from 4.11 to 5.5 min; MG from 11 to 13.6 min; Cer from 13.61 to 15 min; FAs from 15.01 to 17 min; PG from 20.51 to 22 min; PE from 26.01 to 28 min; PI and PS from 28.01 to 30.5 min; PC from 30.51 to 35 min; SM from 35.01 to 39 min; *sn*-1 LPC from 39.6 to 41 min; *sn*-2 LPC from 41.5 to 42.4 min.

RPLC-MS/MS analysis

The RPLC-MS/MS analysis was performed as we reported previously [13], using a Vanquish UHPLC system coupled to a high-resolution hybrid Quadrupole-Orbitrap mass spectrometer (QExactive HF, Thermo Fisher Scientific, USA). The separation was achieved using an Accucore C30 column (Thermo Fisher Scientific) maintained at 40 °C and the gradient was delivered at a flow rate of 350 μL/min. The mobile phases A and B were ACN/H₂O (60:40, v/v) and IPA/ACN (90:10, v/v), respectively, both containing 10 mM NH₄HCO₃ and 0.1% HCOOH [12, 13]. The sample tray was set at 15 °C with the injection volume of 5 μL.

The following parameters were used in electrospray ionization: the spray voltage, the capillary temperature, and the heater temperature were at 3 kV, 350 °C, and 400 °C, respectively, for both ionization modes; the S-Lens RF level was set at 50. The Orbitrap mass analyzer was operated at a resolving power of 120,000 in full-scan mode (scan range, 114–1700 *m/z*; automatic gain control target, 1e⁶) and of 30,000 in the Top20 data-dependent MS2 mode (HCD fragmentation with stepped normalized collision energy: 25 and 30 in positive ion mode and 20, 24, and 28 in negative ion mode; maximum ion injection time, 100 ms; isolation window, 1 *m/z*; automatic gain control target, 1e⁵ with dynamic exclusion setting of 15 s).

Fatty acid analysis

The method for total fatty acid analysis was described in our previous work [12]. Briefly, total lipids were extracted from human serum and dried prior to hydrolysis and derivatization with 2 M KOH/CH₃OH. The resulted fatty acid methyl esters (FAMES) were

extracted and profiled using an Agilent 7890 GC system (Agilent Technologies, Santa Clara, CA) coupled to a Leco Pegasus HT time-of-flight MS (Leco, St. Joseph, MI). A HP-88 column (100 m × 0.25 mm) with a film thickness of 0.2 μm (Agilent Technologies) was utilized for separation of FAMES. Commercial FAMES standard mixture (Agilent) was used as reference standards.

Data processing

All LC-MS/MS data files were processed using the LipidSearch software (version 4.1) (Thermo Fisher Scientific) to identify lipid molecular species within each lipid fraction. Settings of LipidSearch were as follows: precursor tolerance, 5 ppm; product tolerance, 5 ppm; product ion threshold, 5%; m-score threshold, 1; Quan m/z tolerance, ± 5 ppm; Quan RT (retention time) range, ± 0.5 min; use of main isomer filter and ID quality filters A, B, C, and D; adduct ions: +H and +NH₄ for positive ion mode, and -H, +HCOO, and -2H for negative ion mode. The lipid classes selected for the search were LPC, PC, LPE, PE, LPS, PS, LPG, PG, LPI, PI, LPA, PA, SM, MG, DG, TG, CL, So, Cer, and CE. The same lipid annotations identified within ± 0.1 min were merged into the aligned results. LipidSearch results were manually inspected for *sn*-positional assignment of fatty acyls according to corresponding fragment ion intensities in tandem mass spectrometry. Identification results were further filtered using the retention time-total fatty acyl chain length correlation and confident identification was also restricted to fatty acyl compositions provided by GC-MS-based total fatty acid analysis.

The resultant raw data files of the pre-T1D sample were processed using Progenesis QI (Nonlinear Dynamics, UK) for peak detection with the following parameters: peak percentage 0.04%, retention time windows from 1 to 25 min. All detected features were searched against our human serum lipid AMT library with matching tolerances of 0.2 min in retention time and 10 ppm in mass accuracy.

Results

Isomeric and isobaric lipid species are dominant in highly complex blood serum samples, in addition, highly abundant lipid species can mask the low abundant lipid species, and similarly the high ionization efficiency species can interfere with the detection of species with low ionization efficiency. Therefore, to broaden the coverage of serum lipidome and to provide highly confident lipid identification, a mixed-mode-LC coupled with an ELSD detector was used to fractionate the total lipid extract into different lipid classes according to the polarity of lipid head group, which is followed by further separation of the collected fractions into individual molecular species and structural characterization using RPLC-MS/MS analysis (Fig. 1).

Fractionation of total lipid extracts using mixed-mode LC to simplify the sample complexity of downstream RPLC-MS/MS

We applied a previously optimized mixed-mode LCELSD method for class-level separation of lipids from total lipid extract of serum [12]. In total, 18 lipid classes and subclasses were detected and eluted in the order of increasing polarity (Fig. 2), namely, cholesterol ester

(CE), triacylglycerol (TG), cholesterol (Chol), phosphatidylglycerol (PG), phosphatidylethanolamine (PE), phosphatidylethanolamine plasmalogen (p-PE), lysophosphatidylglycerol (LPG), phosphatidylinositol (PI), phosphatidylserine (PS), lysophosphatidylethanolamine (LPE), phosphatidylcholine (PC), phosphatidylcholine plasmalogen (p-PC), sphingomyelin (SM), *sn*2-lysophosphatidylcholine (*sn*2-LPC), and *sn*1-lysophosphatidylcholine (*sn*1-LPC); ceramide (Cer), 1,3-diacylglycerol (1,3-DG), and 1,2-diacylglycerol (1,2-DG) existed only in low level in serum; therefore, they were not labeled on the chromatogram. Because of the low levels of LPG, LPE, and LPI in the serum, they co-eluted with other glycerophospholipids, but this did not pose a challenge in downstream RPLC-MS/MS analysis as the co-eluting classes do not share isobaric species (Fig. 2). Although mixed-mode LC-ELSD analysis can distinguish disease-associated dysregulation of metabolism at the lipid class level, overall, there were no noticeable class-level changes between the disease groups or between the early and late time points of each group.

RPLC-MS/MS analysis of lipid class fractions improved confidence in structural identification of lipid molecular species

With mixed-mode LC separation, the cross-class, isobaric species used to co-elute were separated into different fractions for further downstream RPLC-MS/MS-based molecular level separation and identification. This greatly improved the confidence of structural assignment. For instance, a full-scan MS of PC fraction showed a lipid species with a neutral mass of 755.5465 Da. According to Human Metabolome Data Base (www.hmdb.ca) and LIPID MAPS (www.lipidmaps.org), 32 and 28 isomeric lipid species from different lipid classes are associated with this mass, respectively, each with a unique composition. However, because this species was from the PC fraction, we limited the database search space of LipidSearch to PC when processing the RPLC-MS/MS raw data of this fraction. Specifically, the precursor ion at m/z 756.5529 ($[M+H]^+$) was selected for MS/MS fragmentation in positive ionization mode. As shown in Fig. 3a, the base peak at m/z 184.0733—a signature ion of phosphocholine head group confirmed the identity of the fraction collected in the mixed-mode LC separation as PC. Similarly, m/z 800.5445 ($[M+HCOO]^-$) was selected in the negative ionization mode for MS/MS, and it aligned well in retention time (< 0.1 min) with the positive mode data. As shown in Fig. 3b, prominent ions at m/z 253.2172 and 279.2329, corresponding to the ions of $[FA(16:1)-H]^-$ and $[FA(18:2)-H]^-$, respectively, demonstrated that the fatty acyl composition of this species is 16:1 and 18:2. The m/z 279.2329 as the base peak indicated that FA 18:2 located at *sn*-2 position on the glycerol backbone, as it is known that fragment ion ratio of $[FA_1-H]^-/[FA_2-H]^-$ is less than 1 [13, 16, 27]. Hence, we concluded the identification for the species with neutral mass of 755.5465 Da with the retention time of 9.61 min to be PC 16:1/18:2.

Although fragment ion ratio of $[FA_1-H]^-/[FA_2-H]^- < 1$ has been commonly used to assign the *sn*-positions, there are reported exceptions to this rule and it casts doubt on solely relying on this ratio to assign the *sn*-position when using different collision energies [28–31]. Therefore, we performed MS/MS for glycerolphospholipids and triacylglycerols standards containing different compositions of fatty acyl chains. These standards were studied with the stepped collision energy listed in the method section, under both direct infusion and LC-MS

conditions. In agreement with the literature [13, 29, 32–35], our data suggested that for most of the phospholipids, generation of $[FA_2-H]^-$ is more favorable than $[FA_1-H]^-$ and the ratio of $[FA_1-H]^-/[FA_2-H]^-$ is consistently smaller than 1 at different normalized collision energy. As a result, majority of the lipid species identified in this work have clear annotation of *sn*-positions using the “/”; for lipids with uncertainty in assigning the *sn*-position, we annotated it with “_” [36, 37].

Elution order of lipids on RPLC column depends on the total carbon number and degree of unsaturation of the fatty acyls

Relative retention on RPLC column depends on the hydrophobicity of analyte, the more hydrophobic, the later it elutes. In the case of lipids, this is determined primarily by the length and saturation of fatty acyls. This rule is especially true for the molecular species within the same class of lipids, where all species share the same head group [38, 39]. We plotted the retention time versus the total carbon number of fatty acyls and observed, as shown in Fig. 4, a relationship of second degree polynomial regression for the fully saturated species of the abundant serum lipid classes: TG, SM, PC, LPC, and Cer, with each class having a different slope reflecting the actual gradient conditions experienced by species in each class. In these plots, we observed species with the same number of total carbon but different retention time. By studying their MS/MS profile in both positive and negative modes, we confirmed they are *sn*-position isomers. Nevertheless, the correlation coefficients R^2 are greater than 0.99 in all plots. To our knowledge, this is the most comprehensive determination of the dependence between elution order and total carbon number of various species of different lipid classes.

Similar correlation was observed for the fatty acyls containing one, two, and multiple double bonds (Fig. 5). Compared with lipids containing fully saturated fatty acyls, more lipid molecules with the same total number of carbon and double bond exhibited different retention time (vertically lined dots). These species are isomers with different compositions of fatty acyl chains, in particular the C=C positional isomers, as the interactions are slightly different between different locations of C=C double bond and the C30 RPLC column. As a result of the presence of these isomeric lipid species, the correlation R^2 are slightly lower than the values of the fully saturated lipids in Fig. 4, but still are greater than 0.97 with the species having less than three double bonds. When the lipid species have at least three double bonds, the correlation R^2 were generally reduced, with the SM and TG lipids significantly reduced to > 0.80 . Overall, PC had the highest correlation between retention time and total carbon number, even when the degree of unsaturation increased, while TG and SM had a lower correlation coefficient due to the complexity of the structures and a higher number of isomers (more vertical dots) for each species.

In addition, higher degree of unsaturation resulted in the lipid species eluting earlier when comparing the lipid species with the same total number of carbon and different number of double bonds, i.e. for the same lipid class, the X:2 lipid species eluted earlier than X:1 when X is identical (Fig. 5), which is in agreement with previous report [38]. It is of note that the excellent correlation observed between the retention time on RPLC and the length and degree of unsaturation of fatty acyls provide a way to filter the output of the LipidSearch

software. As a result, those putative lipid identifications that deviated from the established regression curves were further filtered out and not counted as valid.

GC-MS-based fatty acyl analysis improved confidence in identification of fatty acyl composition of lipids

Lipidomics research requires automated data annotation software, particularly at the initial identification stage. The LipidSearch software is highly sensitive in peak detection; hence, often times, species in the noise were considered as identifications. Another observation from using LipidSearch as well as other automated software is the ability to identify “biological impossible” lipids due to the input of the *in silico* library. The large number of putative identifications resulting from combinatorial enumeration inevitably generates large number of false positive identifications. To this end, we performed fatty acyl methyl ester analysis using GC-MS by derivatizing the fatty acyls hydrolyzed from lipid backbone under alkaline methanol condition. Using methyl ester derivatives of fatty acid standards as reference, fatty acyl chains of serum lipids were analyzed. Results showed that the fatty acyl chain lengths varied from 8 to 24 carbons, including the odd numbered chains, and the degree of unsaturation varied from 0 to 6. While it is possible that serum lipids contain unique fatty acyls of odd numbered carbon (> 21 carbons) or more than 6 double bonds, their low abundance in these serum samples did not provide strong enough justification for the inclusion of these fatty acyls as valid lipid identifications. Based on the results of fatty acyl analysis, we filtered out the putative lipid identifications to limit the valid identifications only to those ones containing the detected fatty acyls.

Contents and characteristics of the human serum lipid library

Using the offline 2D-LC-MS/MS approach and the data filtering and validating approaches mentioned above, we confidently identified 753 lipid molecular species in healthy control and pre-T1D human sera. The curated library, as provided in Table S1 (see the Electronic Supplementary Material (ESM)), contains information of accurate mass and RPLC retention time for each identified species. Overall, these 753 lipid molecules belong to 13 major lipid classes: Cer, DG, LPC, LPE, LPI, LPG, TG, PC, PI, PE, PS, PG, and SM (Fig. 6). The subclasses including plasmalogen and *sn*-position (*1,3*- and *1,2*-DG and *sn*-*1*/*sn*-*2* LPC) were combined and reported under their corresponding representative classes such as DG, PC, PE, LPE and LPC. Every lipid class was detected in both positive and negative ionization mode, except TG and Cer, which were solely in positive mode. $[M+H]^+$ and $[M+NH_4]^+$ are the dominant ions in positive ionization mode, while $[M-H]^-$ and $[M+HCOO]^-$ in the negative mode. In total, 88% of the lipids within our human serum lipid library are unsaturated lipid species, with 83% of them being polyunsaturated (degree of unsaturation

2). Majority of the species observed belong to TG, SM, PC, and PE classes. Cer species are composed of fatty acids of 16 to 24 carbons, with the majority having one to two double bonds and an appearance of the odd long chain base d17:1. Both *1,2*-DG and *1,3*-DG were detected, with a predominance of polyunsaturated *1,2*-DG species. *Lys*-glycerophospholipids (LPLs) including LPC, LPI, LPG, and LPE were also found in the serum sample with a diverse fatty acyl chain length of 12–22 carbon. LPC is the most dominant LPLs with the strong preference of *sn*-*1* over *sn*-*2* isomers. Moreover, we were able to clearly identify plasmalogen and plasmalogen PLs using their distinct MS/MS fragments.

We further plotted the spatial distribution of all lipids included in the library in the two-dimensional spaces of RPLC retention time and m/z . As shown in Fig. 7, clusters of lipids with mass ranges from 400 to 600 Da, 500 to 900 Da, and > 900 Da eluted in the following order: earlier than 5 min, 5.5–17 min, and after 17 to 25 min, respectively. Clearly, many isobaric and isomers species overlapped and it would be challenging to identify them if only using RPLC-MS/MS. The only lipid class that was separated well from others on RPLC is the highly non-polar TG. Lipid species belonging to LPLs eluted together, and GPLs (PC, PE, PI, PS) co-eluted with the DG, SM, and Cer species. However, DG and Cer species have a different mass range from phospholipids. Conversely, with mixed-mode LC, we increased the level of confidence in lipid identification because a correct identification needs to come from the right fraction (lipid class) and matches with the accurate mass and the elution order.

Application of the lipid library to identify lipids from serum sample

As a demonstration of the utility of the curated library in identification of lipids from biological sample, total lipids extracted from one individual pre-T1D serum sample were directly analyzed by RPLC-MS/MS without fractionation at lipid class level and raw data files were processed using two approaches—without and with our AMT serum lipids database. In the first approach, the raw data were processed directly using the LipidSearch software resulting into 217 lipid species identified with assignment of *sn*-positions. In the second approach, the resultant raw data were processed using the Progenesis QI software. Detected features from Progenesis QI were searched against the curated human serum lipid library based on the accurate mass and RPLC retention time of each lipid molecular species. In total, we identified 412 molecular lipids in one individual pre-T1D serum sample, as summarized in Fig. 8. Majority of the species identified are in the class of TG, PC, SM, PE, and LPC, which is proportional to the composition of lipid library. The profile of carbon length and degree of unsaturation for each fatty acyl in T1D lipidome varied from 10 to 22 carbons, including some odd chains (15 and 17 carbons) and 0–6 double bonds C=C. Hence, with the use of our AMT lipid serum database, we are able to increase the number of identification significantly, mostly from the low abundant species.

Discussion

There has been an increased interest in comprehensively characterizing the human serum lipidome for its use in disease diagnosis. However, due to the complexity of blood serum samples, it is commonly recognized that many lipid species are unlikely to be detected due to their low abundance in serum, which is further hindered by the limited resolving power of one-dimensional LC separation and the under-sampling issue of data-dependent MS/MS. The accurate mass and time (AMT) tag approach was initially developed for high-throughput proteomics to provide extensive coverage of complex peptide mixtures by taking advantage of the high resolution and wide dynamic range of the MS scans, which largely overcame the under-sampling issue in MS/MS-based peptide identification [22]. The concept employed in AMT tag-based proteomics studies could be applied to study other biomolecules, such as lipids [24]. Specifically, if an AMT tag library can be generated for various lipid species using a reproducible and robust LC-MS platform, then lipids from different samples of the same type of specimen could be analyzed in high throughput using

the identical platform, where the AMT tag library would serve as a look-up table for lipid identification using LC retention time and accurate mass instead of relying on the MS/MS spectrum.

To achieve this, we employed offline two-dimensional LC separation [12], which used mixed-mode LC in the first dimension to fractionate the total lipid extract into lipid classes, followed by reversed-phase LC separation to further separate lipids in the same class into molecular species. The chromatographic separation and mass spectrometric detection conditions have previously been optimized in our laboratory, which doubled the coverage of rat serum and liver lipidome compared to using RPLC-MS/MS alone. Fractionating total lipid extract into different lipid classes prior to the second dimensional RPLC-MS/MS analysis is critical for building the library, as it greatly increased the confidence in lipid identification process by limiting the cross contamination of co-eluting isobaric/isomeric species between different lipid classes. For example, isobaric PC and SM can be challenging to differentiate when co-eluted on RPLC column because they share the common characteristic fragment ion (phosphocholine head group) in positive ionization mode [40]. However, the ambiguities in identification can be eliminated with additional dimension of separation provided by mixed-mode LC on which SM is clearly separated from PC.

Automated lipid identification programs, such as LipidSearch, are powerful in identification of lipid molecular species based on pre-configured fragmentation rules [18]. However, not all MS/MS spectra acquired from real biological samples are of high quality, which can be further complicated by lipids co-eluted and co-fragmented within the precursor ion selection window. This undoubtedly poses a challenge in structural elucidation when only head group-specific ions can be observed. As an effort to improve the confidence of identified lipids, we did total fatty acyl analysis using GC-MS to limit the composition of identified lipids only to those detected fatty acyls. We also observed a rigorous dependence between the RPLC retention time and the sum composition of total carbon number of all fatty acyls, which was used to further remove the putative lipid identifications not following the trend line.

The ratio of $[FA_1-H]^-/[FA_2-H]^-$ has been used for *sn*-positional assignments; however, relative intensities of the two carboxylate anions can vary depending on length and unsaturation, as well as the collision energy used in fragmentation [27, 31, 41, 42]. In this respect, it has been reported that different fragmentation patterns originated from different cleavage sites could be formed as a result of very different collision energies [43], or the ratio of $[FA_1-H]^-/[FA_2-H]^-$ could be altered with the collision energy [29, 31]. Hence, to confidently identify the *sn*-position assignment of fatty acyls, we studied the relative ratio of $[FA_1-H]^-/[FA_2-H]^-$ using a list of representative lipid standards with different fatty acyl chains for each class, and concluded that $[FA_1-H]^-/[FA_2-H]^- < 1$ can be used to assign the fatty acyl positions in most of lipid classes under our experimental conditions, i.e., NCE 20–30 under HCD. Thus, within the stepped collision energies used in this work, we ruled out their effects on reversing the ratio of carboxylate anions.

Results from mixed-mode LC separation showed that there were no noticeable lipid class-level changes between healthy and T1D groups or between the early and late time points of each group. Considering the samples used in this study were collected from patients older

than 8 months and pooled from 25 subjects and two time points, our observation is somewhat in agreement with the serum lipid profile changes reported for the Finnish Type 1 Diabetes Prediction and Prevention cohort (DIPP) study, where the lipidomic profile changes associated with the progression of T1D appear to be most pronounced in children at 3 months of age [6]. Nevertheless, no change at the lipid class level further underline the importance of separating lipids at the molecular species level.

In agreement with the literature, our finding revealed that TG, PC, and SM are predominant lipids in human serum, with TG comprising over 50% of lipids having more than 52 fatty acyl carbon atoms [44, 45]. Compared to the previous published human plasma/serum lipidome, our lipid library provided unambiguous identifications at the level of *sn*-assignment for most of the glycerolphospholipids [45–47]. Although generally not native to mammalian cell lipidome, glycerolphospholipids containing odd chain fatty acyls and longer chain fatty acyls on the *sn*-1 position were also observed and passed our validation criteria. It is of note that, despite cholesterol and fatty acids being abundant classes observed in the mixed-mode LC (Fig. 2), we did not perform identification of these two classes at the molecular level. Cholesterols have been known to ionize poorly with electrospray ionization unless derivatized to cholesterol esters [48], and therefore, traditionally cholesterols are ionized with chemical or photoionization techniques and not suitable for detection using our electrospray ionization-based platform. With respect to identification of free fatty acids, the main fragment ions resulted from CO₂ loss requires NCE > 50 [49], which greatly exceed the NCE used for other lipid classes.

In summary, using orthogonal and highly resolving separation methods of mixed-mode LC and RPLC in conjunction with high-resolution tandem mass spectrometry and multiple levels of data filtering and validation, we have created a comprehensive human serum lipid library containing 753 lipid molecular species, with accurate mass and retention time annotated for each lipid molecule and with confident assignment of fatty acyl *sn*-positions for most of the species. This library not only provides a comprehensive resource for studies of T1D, it will also be valuable for biomarker studies of other childhood diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

The work was partially supported by National Institutes of Health (NIH) grants R21 GM104678 and R01 DK114345. Clinical sample collection was supported by NIH grant R01 DK32493.

Biographies

Ngoc Vu is a PhD student at the University of North Carolina at Greensboro, USA. She is working on the development of hyphenated mass spectrometry technologies for structure elucidation of lipids and applications of high-throughput lipidomics to biomedical research.

Monica Narvaez-Rivas was a postdoctoral researcher at the University of North Carolina at Greensboro and currently is a scientist at the mass spectrometry lab in the Pathology Department of the Cincinnati Children's Hospital, USA. She is working in bile acid, lipid, and oxysterol analysis as biomarkers for different diseases such as cystic fibrosis, liver disease, and malnutrition in children, using LC-MS and GC-MS techniques.

Guan-Yuan Chen is a research scientist at the Center for Translational Biomedical Research, University of North Carolina at Greensboro, USA. His expertise is in the development of mass spectrometry-based analytical methods for determining drugs, metabolites, and lipids in various biological matrices and in applying of bioanalytical chemistry to biomedical research.

Marian Rewers is Professor of Pediatrics and Medicine and Executive Director of Barbara Davis Center for Diabetes at the University of Colorado School of Medicine, USA. He is a pediatric endocrinologist who has dedicated his research to finding the cause and prevention of type 1 diabetes (T1D) and its complications through multiple NIH-funded cohort studies: Diabetes Autoimmunity Study in the Young (DAISY), The Environmental Determinants of

Diabetes in the Young (TEDDY), and The Coronary Artery Calcification in Type 1 (CACTI).

Qibin Zhang is Associate Professor in the Department of Chemistry and Biochemistry and Co-director of the Center for Translational Biomedical Research at the University of North Carolina at Greensboro, USA. His research interests are in the development of advanced bioanalytical measurement capabilities and applications of lipidomics, metabolomics, and proteomics to human disease research, with the aim of early diagnosing and better understanding the pathogenic mechanism of diseases.

References

1. Wymann MP, Schneider R. Lipid signalling in disease. *Nat Rev Mol Cell Biol.* 2008;9(2):162–76. [PubMed: 18216772]
2. Nguyen A, Rudge SA, Zhang Q, Wakelam MJO. Using lipidomics analysis to determine signalling and metabolic changes in cells. *Curr Opin Biotechnol.* 2017;43:96–103. [PubMed: 27816901]
3. Li X, Xu Z, Lu X, Yang X, Yin P, Kong H, et al. Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for metabolomics: biomarker discovery for diabetes mellitus. *Anal Chim Acta.* 2009;633(2):257–62. [PubMed: 19166731]
4. Zarrouk A, Debbabi M, Bezine M, Karym EM, Badreddine A, Rouaud O, et al. Lipid biomarkers in Alzheimer's disease. *Curr Alzheimer Res.* 2018;15(4):303–12. [PubMed: 28474568]
5. Fox TE, Bewley MC, Unrath KA, Pedersen MM, Anderson RE, Jung DY, et al. Circulating sphingolipid biomarkers in models of type 1 diabetes. *J Lipid Res.* 2011;52(3):509–17. [PubMed: 21068007]
6. Lamichane S, Ahonen L, Dyrland TS, Kempainen E, Siljander H, Hyoty H, et al. Dynamics of plasma lipidome in progression to islet autoimmunity and type 1 diabetes—Type 1 Diabetes Prediction and Prevention Study (DIPP). *Sci Rep.* 2018;8(1):10635. [PubMed: 30006587]
7. Sorensen CM, Ding J, Zhang Q, Alquier T, Zhao R, Mueller PW, et al. Perturbations in the lipid profile of individuals with newly diagnosed type 1 diabetes mellitus: lipidomics analysis of a Diabetes Antibody Standardization Program sample subset. *Clin Biochem.* 2010;43(12):948–56. [PubMed: 20519132]
8. Lippa PB, Bietenbeck A, Beaudoin C, Giannetti A. Clinically relevant analytical techniques, organizational concepts for application and future perspectives of point-of-care testing. *Biotechnol Adv.* 2016;34(3):139–60. [PubMed: 26808189]
9. Wenk MR. The emerging field of lipidomics. *Nat Rev Drug Discov.* 2005;4(7):594–610. [PubMed: 16052242]
10. Holcapek M, Jirasko R, Lisa M. Recent developments in liquid chromatography-mass spectrometry and related techniques. *J Chromatogr A.* 2012;1259:3–15. [PubMed: 22959775]
11. Cajka T, Fiehn O. Increasing lipidomic coverage by selecting optimal mobile-phase modifiers in LC–MS of blood plasma. *Metabolomics.* 2016;12(2).
12. Narvaez-Rivas M, Vu N, Chen GY, Zhang Q. Off-line mixed-mode liquid chromatography coupled with reversed phase high performance liquid chromatography-high resolution mass spectrometry to improve coverage in lipidomics analysis. *Anal Chim Acta.* 2017;954:140–50. [PubMed: 28081809]

13. Narvaez-Rivas M, Zhang Q. Comprehensive untargeted lipidomic analysis using core-shell C30 particle column and high field Orbitrap mass spectrometer. *J Chromatogr A*. 2016;1440:123–34. [PubMed: 26928874]
14. Gao X, Zhang Q, Meng D, Isaac G, Zhao R, Fillmore TL, et al. A reversed-phase capillary ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) method for comprehensive top-down/bottom-up lipid profiling. *Anal Bioanal Chem*. 2012;402(9):2923–33. [PubMed: 22354571]
15. Rampler E, Criscuolo A, Zeller M, El Abiead Y, Schoeny H, Hermann G, et al. A novel lipidomics workflow for improved human plasma identification and quantification using RPLC-MSn methods and isotope dilution strategies. *Anal Chem*. 2018;90(11):6494–501. [PubMed: 29708737]
16. Wozny K, Lehmann WD, Wozny M, Akbulut BS, Brugger B. A method for the quantitative determination of glycerophospholipid regioisomers by UPLC-ESI-MS/MS. *Anal Bioanal Chem*. 2019;411(4):915–24. [PubMed: 30580388]
17. Kind T, Liu KH, Lee DY, DeFelice B, Meissen JK, Fiehn O. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat Methods*. 2013;10(8):755–8. [PubMed: 23817071]
18. Taguchi R, Ishikawa M. Precise and global identification of phospholipid molecular species by an Orbitrap mass spectrometer and automated search engine LipidSearch. *J Chromatogr A*. 2010;1217(25):4229–39. [PubMed: 20452604]
19. Meng D, Zhang Q, Gao X, Wu S, Lin G. LipidMiner: a software for automated identification and quantification of lipids from multiple liquid chromatography/mass spectrometry data files. *Rapid Commun Mass Spectrom*. 2014;28(8):981–5. [PubMed: 24623704]
20. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*. 2010;11:395. [PubMed: 20650010]
21. Conrads TP, Anderson GA, Veenstra TD, Paša-Toli L, Smith RD. Utility of accurate mass tags for proteome-wide protein identification. *Anal Chem*. 2000;72(14):3349–54. [PubMed: 10939410]
22. Zimmer JS, Monroe ME, Qian WJ, Smith RD. Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom Rev*. 2006;25(3):450–82. [PubMed: 16429408]
23. Creek DJ, Jankevics A, Breitling R, Watson DG, Barrett MP, Burgess KE. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. *Anal Chem*. 2011;83(22):8703–10. [PubMed: 21928819]
24. Ding J, Sorensen CM, Jaitly N, Jiang H, Orton DJ, Monroe ME, et al. Application of the accurate mass and time tag approach in studies of the human blood lipidome. *J Chromatogr B Anal Technol Biomed Life Sci*. 2008;871(2):243–52.
25. Rewers M, Bugawan TL, Norris JM, Blair A, Beaty B, Hoffman M, et al. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia*. 1996;39(7):807–12. [PubMed: 8817105]
26. Norris JM, Yin X, Lamb MM, Barriga K, Seifert J, Hoffman M, et al. Omega-3 polyunsaturated fatty acid intake and islet autoimmunity in children at increased risk for type 1 diabetes. *JAMA*. 2007;298(12):1420–8. [PubMed: 17895458]
27. Nakanishi H, Iida Y, Shimizu T, Taguchi R. Separation and quantification of sn-1 and sn-2 fatty acid positional isomers in phosphatidylcholine by RPLC-ESIMS/MS. *J Biochem*. 2010;147(2):245–56. [PubMed: 19880374]
28. Larsen A, Uran S, Jacobsen PB, Skotland T. Collision-induced dissociation of glycerophospholipids using electrospray ion-trap mass spectrometry. *Rapid Commun Mass Spectrom*. 2001;15(24):2393–8. [PubMed: 11746908]
29. Hvattum E, Hagelin G, Larsen A. Study of mechanisms involved in the collision-induced dissociation of carboxylate anions from glycerophospholipids using negative ion electrospray tandem quadrupole mass spectrometry. *Rapid Commun Mass Spectrom*. 1998;12(19):1405–9. [PubMed: 9773525]

30. Cui Z, Thomas MJ. Phospholipid profiling by tandem mass spectrometry. *J Chromatogr B Anal Technol Biomed Life Sci*. 2009;877(26):2709–15.
31. Pi J, Wu X, Feng Y. Fragmentation patterns of five types of phospholipids by ultra-high-performance liquid chromatography electrospray ionization quadrupole time-of-flight tandem mass spectrometry. *Anal Methods*. 2016;8(6):1319–32.
32. Hsu FF, Turk J. Structural characterization of unsaturated glycerophospholipids by multiple-stage linear ion-trap mass spectrometry with electrospray ionization. *J Am Soc Mass Spectrom*. 2008;19(11):1681–91. [PubMed: 18771936]
33. Hsu FF, Turk J. Electrospray ionization multiple-stage linear ion-trap mass spectrometry for structural elucidation of triacylglycerols: assignment of fatty acyl groups on the glycerol backbone and location of double bonds. *J Am Soc Mass Spectrom*. 2010;21(4):657–69. [PubMed: 20171120]
34. Hsu FF, Turk J, Thukkani AK, Messner MC, Wildsmith KR, Ford DA. Characterization of alkylacyl, alk-1-enylacyl and lyso sub-classes of glycerophosphocholine by tandem quadrupole mass spectrometry with electrospray ionization. *J Mass Spectrom*. 2003;38(7):752–63. [PubMed: 12898655]
35. Han X, Yang K, Gross RW. Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. *Mass Spectrom Rev*. 2012;31(1):134–78. [PubMed: 21755525]
36. Holapek M, Liebisch G, Ekroos K. Lipidomic analysis. *Anal Chem*. 2018;90(7):4249–57. [PubMed: 29543437]
37. Liebisch G, Vizcaino JA, Kofeler H, Trotzmüller M, Griffiths WJ, Schmitz G, et al. Shorthand notation for lipid structures derived from mass spectrometry. *J Lipid Res*. 2013;54(6):1523–30. [PubMed: 23549332]
38. Ovcacikova M, Lisa M, Cifkova E, Holcapek M. Retention behavior of lipids in reversed-phase ultrahigh-performance liquid chromatography-electrospray ionization mass spectrometry. *J Chromatogr A*. 2016;1450:76–85. [PubMed: 27179677]
39. Nakagawa Y, Horrocks LA. Separation of alkenylacyl, alkylacyl, and diacyl analogues and their molecular species by high performance liquid chromatography. *J Lipid Res*. 1983;24(9):1268–75. [PubMed: 6631251]
40. Merrill AH Jr, Sullards MC, Allegood JC, Kelly S, Wang E. Sphingolipidomics: high-throughput, structure-specific, and quantitative analysis of sphingolipids by liquid chromatography tandem mass spectrometry. *Methods*. 2005;36(2):207–24. [PubMed: 15894491]
41. Smith PB, Snyder AP, Harden CS. Characterization of bacterial phospholipids by electrospray ionization tandem mass spectrometry. *Anal Chem*. 1995;67(11):1824–30. [PubMed: 9306733]
42. Han X, Gross RW. Structural determination of picomole amounts of phospholipids via electrospray ionization tandem mass spectrometry. *J Am Soc Mass Spectrom*. 1995;6(12):1202–10. [PubMed: 24214071]
43. Bird SS, Marur VR, Sniatynski MJ, Greenberg HK, Kristal BS. Serum lipidomics profiling using LC-MS and high-energy collisional dissociation fragmentation: focus on triglyceride detection and characterization. *Anal Chem*. 2011;83(17):6648–57. [PubMed: 21774539]
44. Quehenberger O, Dennis EA. The human plasma lipidome. *N Engl J Med*. 2011;365(19):1812–23. [PubMed: 22070478]
45. Quehenberger O, Armando AM, Brown AH, Milne SB, Myers DS, Merrill AH, et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. *J Lipid Res*. 2010;51(11):3299–305. [PubMed: 20671299]
46. Zhou X, Mao J, Ai J, Deng Y, Roth MR, Pound C, et al. Identification of plasma lipid biomarkers for prostate cancer by lipidomics and bioinformatics. *PLoS One*. 2012;7(11):e48889. [PubMed: 23152813]
47. Sandra K, Pereira Ados S, Vanhoenacker G, David F, Sandra P. Comprehensive blood plasma lipidomics by liquid chromatography/quadrupole time-of-flight mass spectrometry. *J Chromatogr A*. 2010;1217(25):4087–99. [PubMed: 20307888]

48. Liebisch G, Binder M, Schifferer R, Langmann T, Schulz B, Schmitz G. High throughput quantification of cholesterol and cholesteryl ester by electrospray ionization tandem mass spectrometry (ESI-MS/MS). *Biochim Biophys Acta*. 2006;1761(1):121–8. [PubMed: 16458590]
49. Murphy RC, Axelsen PH. Mass spectrometric analysis of long-chain lipids. *Mass Spectrom Rev*. 2011;30(4):579–99. [PubMed: 21656842]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.

Workflow to create the accurate mass and time tag library for serum lipids. Total lipids were extracted from pooled samples. In the first dimension of LC separation, fractions containing lipid classes from total lipid extraction were separated and collected using mixed-mode LC-ELSD and further analyzed at the molecular level in the second LC dimension using RPLC-MS/MS. Putative identifications obtained from the automated data processing software LipidSearch were manually validated using multiple data filtering criteria (total fatty acid analysis by GC-MS, MS/MS profile, and LC elution order), only the verified lipid identifications were curated into the final lipid library with each species annotated with accurate mass and RPLC retention time

Fig. 2. Mixed-mode LC-ELSD chromatograms obtained from total lipid extract of human sera. **a** Pre-T1D samples and **b** healthy control sample. Both chromatograms showed similar lipid profiles at lipid class level. 1: CE, 2: TG, 3: Chol, 4: FA, 5: PG, 6: LPG, PE, p-PE, 7: PI, PS, LPE, 8: PC, p-PC, LPI, 9: SM, 10: *sn2*-LPC, *sn1*-LPC

Fig. 3. MS/MS spectra of PC 16:1/18:2 in **a** positive ion mode and **b** negative ion mode. m/z 184.0733 in **a** shows the signature ion of PC, which also confirmed the accuracy of mixed-mode LC fractionation, while m/z 253.2172 and 279.2329 in **b** reflect the composition of the two fatty acyl chains and the *sn*-position assignment

Fig. 4.

Plots showing dependences of RPLC retention times of saturated lipids to the total number of carbon; each panel is a different lipid class (TG, SM, PC, Cer, and LPC). Lipid molecules with longer fatty acyl chain (larger number of total carbon) are less polar and elute later from column

Fig. 5.

Plots showing dependences of RPLC retention times of unsaturated lipids to the total numbers of carbon, with each panel showing a different lipid class (TG, SM, and PC) or degree of unsaturation (1 to 4). Within each class, slightly different retention times were observed between different C=C positional isomers and *sn*-position isomers (vertically lined dots)

Fig. 6.
Relative distribution of lipid species contained in the serum lipid library

Fig. 7.

Individual lipid molecular species identified in human sera, plotted with their respective m/z and retention time (min) on the RPLC. Identification of these lipid molecular species was facilitated by lipid class level fractionation using mixed-mode LCELSD and molecular species level separation using RPLC-MS/MS.

Fig. 8.

Number of lipid species identified from a T1D serum sample. Progenesis QI was used to match the LC-MS profiles of the T1D serum sample to the serum lipid AMT library