Breast milk metabolome characterization in a single phase extraction, multiplatform analytical approach

Alma Villaseñor[1#], Isabel Garcia-Perez[2, 3#], Antonia Garcia[1], Joram M. Posma[2], Mariano Fernández-López[4], Andreas J Nicholas[5], Neena Modi[5], Elaine Holmes[2], Coral Barbas[1*].

[1]Centre for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy, Universidad San Pablo CEU, Campus Monteprincipe, Boadilla del Monte, 28668 Madrid, Spain.

[2]Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ London, UK.

[3]Nutrition and Dietetic Research Group, Division of Endocrinology and Metabolism, Faculty of Medicine, Imperial College London, W12 0NN, UK.

[4]Escuela Politécnica Superior. Universidad San Pablo CEU, Campus Monteprincipe, Boadilla del Monte, 28668 Madrid, Spain.

[5]Section of Neonatal Medicine, Department of Medicine, Chelsea & Westminster Hospital campus, Imperial College London, London, SW10 9NH, UK.

ABSTRACT: Breast milk (BM) is a biofluid, which has a fundamental role in early-life nutrition and directly impacts on growth, neurodevelopment and health. Global metabolic profiling is increasingly being utilized to characterize complex metabolic changes in biological samples. However, in order to achieve broad metabolite coverage, it is necessary to employ more than one analytical platform, typically requiring multiple sample preparation protocols. In an effort to improve analytical efficiency and retain comprehensive coverage of the metabolome, a new extraction methodology was developed that successfully retains metabolites from BM in a single-phase using an optimized methyl-*tert*-butyl ether solvent system. We conducted this single-phase extraction procedure on a representative pool of BM, and characterized the metabolic composition using LC-QTOF-MS and GC-Q-MS for polar and lipidic metabolites. To ensure that the extraction method was reproducible and fit-for-purpose, the analytical procedure was evaluated on both platforms using 18 metabolites selected to cover a range of chromatographic retention times and biochemical classes. Having validated the method, the metabolic signature of BM composition was mapped as a metabolic reaction network highlighting interconnected biological pathways and showing that the LC-MS and GC-MS platforms targeted largely different domains of the network. Subsequently, the same protocol was applied to ascertain compositional differences between BM at week 1 (n=10) and 4 weeks (n=9) post-partum. This single-phase approach is more efficient in terms of time, simplicity, cost and sample volume than the existing two phase methods, and will be suited to high-throughput metabolic profiling studies of BM.

INTRODUCTION

Breast milk (BM) is considered the optimal nutrition available for neonates and is recommended as the sole source of nutrition in the first 6 months of infancy by the World Health Organization. BM contains many biologically active components, which have a fundamental role in infant growth and health[1,2]. In addition to the direct nutritional value of BM in promoting growth, bioactive components found in BM contribute to the metabolic, immune and gut microbiome development of the newborn[3].

The major groups of nutrients present in BM are lipids, carbohydrates, proteins and many other micronutrients in smaller quantities[1,4]. BM is a compositionally variable fluid that changes during lactation, and is highly influenced by maternal dietary intake[5,6]. Preclinical findings show that qualitative changes in dietary lipids during early life may contribute to long-term health outcomes, such as lower neurodevelopmental and maturation of the digestive function[7,8]. Previous studies[4,6] have shown the influence of maternal nourishment on BM lipid content. Jensen, et. al. described temporal differences in lipid composition, especially fatty acid content, over time postpartum, compositional variation from mothers with different ethnicities and diets, and between term and preterm milk[4].

Milk is a complex biological fluid where fat is stabilized in the water phase by micelles due to the presence of several compounds acting as surfactants. After freezing, fat is agglutinated and the sample separates into two layers that are not stably mixed again after stirring or vortexing. Therefore, most published studies have adopted a two phase strategy and have analyzed the aqueous and organic phases separately. Given the complexity of BM, and its importance in human growth, there is a need for development of global metabolic profiling methods that are capable of capturing as

many nutrients and micronutrients present in the BM to give a comprehensive coverage of the metabolome with a view to establishing the influence of the chemical interaction between mother and infant via BM.

Over recent years there has been extensive research into unraveling the metabolome of various biofluids[9]. The main aims of any global metabolic screening method is to achieve broad and rich coverage of as wide a range of endogenous and exogenous compounds in a single and robust profile that represents the physiological or pathological condition of the subject studied[10]. Widespread analytical methodologies such as NMR spectroscopy and more recently GC, LC, and CE coupled to MS have proved valuable in the global assessment of metabolites in biofluids[11-13]. Despite the importance of BM in human development, only two global metabolic profiling studies are reported in the literature[14,15], both using ^{1}H NMR as the analytical platform. In addition, most of the analytical research in BM to date has been targeted and has focused primarily on the fatty acid content using liquid and gas chromatography[16-19]. Also other specific molecular classes such as the oligosaccharides[20] and sphingomyelins[21] have been profiled; for example, Miller, et. al.[20] have shown complex changes in oligosaccharide structure and composition that are unique to mother-infant pairs and that the oligosaccharide composition directly effects the composition of the infant gut microbiome.

Apart from two studies using NMR[14,15] spectroscopy, to the best of our knowledge, no global MS-based assays have been used to study the global profile of BM. A robust screening method for BM capturing multiple classes of metabolites is necessary in order to carry out effective, high throughput screening of this biofluid. The aim of this study was to develop and optimize a methodology for analyzing BM that captures as many metabolites as possible using a single extraction phase in order to maximize the likelihood of identifying biomarkers associated with infant development and health.

We developed a single optimized organic phase extraction method for breast milk suitable for both GC-MS and LC-MS and applied this method to the characterization of differences in BM composition at two distinct times post-partum.

MATERIALS AND METHODS

Reagents

All standards, chemicals, reagents and solvents used in this study are fully described in the Supplementary Information (SI) and compounds used for the validation of both MS platforms are summarized in Table S-1.

Samples

This study was approved by the National Research Ethics Service (ref 12/LO/0203). A pool of human breast milk was obtained mixing 100 µL aliquots from samples obtained from 52 women ranging from 1 to 76 days post-partum in order to characterize BM over the first four months of life. In addition, BM samples from mothers of healthy term-born infants were collected in the 1st week (n=10) and in week 4 post-partum (n=9), and were analyzed in order to show the applicability of the method characterizing differences in BM composition over time. These two time points were chosen since the first week of neonatal nutrition is markedly different from subsequent weeks with colostrum being produced over the first 3-4 days post-partum. This application could be important to prepare special formulas for newborns more similar to the stages of breastfeeding. All samples were stored at -80°C for long term storage and -20°C for short term storage before the day of analysis.

Sample extraction protocol

Several extraction mixtures were compared in order to optimize the extraction protocol for as many metabolites as possible from the BM sample giving coverage of both lipids and polar metabolite classes.

Methanol-MTBE and ethanol-MTBE systems. 50 µL of BM was mixed with 350 µL of the solvents, methanol and MTBE in different ratios: methanol (0, 87, 175, 263, and 350 µL) and the converse volume of MTBE (350, 263, 175, 87, and 0 µL) were tested. The mixture then was vortex for 1 min for protein precipitation and compound extraction. The same protocol was followed for the mixture ethanol: MTBE.

In all cases vitamin E acetate was used as internal standard (IS), at final concentration of 25 ppm.

GC-MS sample treatment. Once the extraction protocol was optimized and established, analysis of the single phase extraction mixture by GC-Q-MS was performed. The analytical procedure for metabolic characterization was carried out following a methodology previously described[22]. Briefly, an aliquot of 150 µL was transferred to a GC vial taken from the supernatant of the extracting system. The sample was evaporated to dryness in a Speedvac Concentrator (Thermo Fisher Scientific, Waltham, MA, USA). 10µL of O-methoxyamine hydrochloride in pyridine (15 mg/mL) was added to each GC vial, and the resultant mixture was vigorously vortex-mixed for 5 minutes. Methoxymation was carried out in the dark, at room temperature for 16 h. For derivatization, the solution was vortex-mixed again for 5 min after 20µL of BSTFA with 1% TMCS was added as catalyst. Samples were heated in an oven for 1 h at 70 °C to achieve silylation. Finally 100 µL of heptane containing 10 ppm of C18:0 methyl ester (IS) was added to each GC vial and vortex-mixed for 2 min before GC analysis (labeled sample).

A flowchart of the extraction protocol for the optimized method is presented in Figure 1.

LC-MS analysis method. Global profiling was developed using a LC-QTOF-MS instrument following methods described elsewhere[23]. The LC system consisted of a degasser, binary pump, and autosampler (1200 series, Agilent). 1 µL of extracted sample was injected into a reversed-phase column at 60°C (Agilent; Poroshell EC-C8, 15 cm × 2.1 mm, 2.7 µm) with a pre-column (Supelco; Ascentis Express C8, 0.5 cm× 2.1 mm, 2.7 µm). The gradient consisted of mobile phase A (10 mM ammonium formate in water) and mobile phase B (10 mM ammonium formate in methanol) pumped at 0.5 mL/min within a total run time of 60 min. The gradient started at 75% B, increasing to 96% B in 23 min, then held until 45 min and increased to 100% B by 46 min, and held until 50 min. Starting conditions were returned by 51 min, and a 9 min re-equilibration time was included taking the total run time to 60 min.

An online filter was added to the instrumentation. Data were collected in positive and negative ESI mode in separate runs on a QTOF analyzer (Agilent 6520) operated in full scan mode from 100 to 1200 m/z. The capillary voltage was 3500 V for positive and 4000 V for negative with a scan rate of 1.02 scans per second; the nebulizer gas flow rate was 10 L/min. During the experiment two reference masses were continuously detected allowing constant mass correction of data to obtain accurate mass in all injections [$(M_1\text{-}H)^+$= 121.0509 ($C_5H_4N_4$) and $(M_2\text{-}H)^+$= 922.0098 ($C_{18}H_{18}O_6N_3P_3F_{24}$)].

GC-MS analysis method. Labeled samples were analyzed using a GC instrument (Agilent 7890A) coupled to mass spectrometer with triple-Axic detector (Agilent 5975C). 2 μL of sample volume was injected in split mode using an Agilent 7693 autosampler. The split ratio was 1:5 to 1:10 with 3–10 mL/min helium split flow into a Restek 20782 deactivated glass wool split liner. Separation was achieved using a 10 m J&W precolumn (Agilent Technologies®) integrated with a 122–5332G column DB5-MS: 30m length, 0.25 mm i.d. and 0.25μm film consisted of 95% dimethyl/5% diphenyl polysiloxane (Agilent Technologies®). The helium carrier gas was used at a constant flow rate of 1 mL/min. The column was initially maintained at 60 °C for 1 min after injection, then temperature was increased at the rate of 10 °C/min to reach a final temperature of 325 °C, and cool down after analysis for 10 min. Temperatures of the injector, transfer line, filament source and quadruple were maintained at 250°C, 280°C, 230°C and 150°C, respectively. The quadrupole detector (5975 inert MSD, Agilent) was controlled by ChemStation software. The electron ionization source was operated at −70 eV. MS detection was achieved in scan mode over a mass range m/z 50–650 at a rate of 1 spectra/s. Filament of the EI source was turned off from minute 23 to 26. An n-alkane mixture from C8 to C28 dissolved in dichloromethane was run at the beginning of experimental worklist for calibration file and retention index determination. All these conditions were optimized previously[22].

For both MS techniques, the samples from the application were analyzed in one randomized run. QC samples were injected in the beginning and at the end of the sequence, and after every 5 samples to check system stability and performance of the analysis.

Data Treatment

LC-MS. The resulting data file was cleaned of background noise and molecular entities were output by the Molecular Feature Extraction (MFE) algorithm in the MassHunter Qualitative Analysis Software (Agilent®, version B.05.00). The resulting MFE file contained a list of all molecular entities that, represents the full TOF mass spectral data in each sample. Primary data treatment (alignment and filtering) was performed in Mass Profiler Professional (MPP) B.12.1 Agilent® software.

GC-MS. Data were acquired with the Agilent MSD ChemStation Software. Identification of metabolites present in the GC-MS profiles was completed before the multivariate analysis. The quality of the chromatograms acquired by Total Ion Chromatogram (TIC) and internal standard peak were carefully examined. Automated Mass Spectrometry Deconvolution and Identification System (AMDIS) was used for automatic peak detection and deconvolution (see compound identification). Compounds were identified by comparing their mass fragmentation patterns with those available in the NIST08 and Fiehn RTL mass spectral library.

Target analysis of 18 compounds was performed. These compounds were used to validate the method. For LC-MS, the Extracted Ion Chromatogram (EIC) algorithm in MassHunter was used; this takes the *m/z* value of each compound and look for this into the chromatogram (±10ppm of mass error window). For GC-MS, extraction and integration of a target fragment ion in MSD ChemStation Software was performed.

Data obtained from the analysis of 19 breast milk samples, obtained at two different time points, were processed using multivariate pattern recognition techniques using SIMCA-P13 software (Umetrics, Sweden) such as principal component analysis (PCA), and partial least square discriminant analysis (PLS-DA)[24]. Unit variance scaling was applied to all data variables. The robustness of the models was evaluated based on R^2 (explained variance) and Q^2 (capability of prediction) values as well as 7-fold cross validation and class permutation validation. The regression coefficients from the PLS-DA models were divided by the jack-knife interval standard error (SE) to give an estimate of the t-statistic. Variables with a |t-statistic| ≥ 1.96 (z-score, corresponding to the 97.5 percentile) were considered significant[25].

Compound Identification

LC-MS. Tentative identification of metabolites was done by comparing mass accuracy of candidate metabolites against our own university online database CEU-mass mediator (http://ceumass.eps.uspceu.es/mediator/), which uses KEGG, METLIN and LipidMAPS databases with an error mass set to 10 ppm. Potential hits were compared to the experimental isotopic pattern distribution on MassHunter.

GC-MS. A standard solution (Grain FAME mix®) was analyzed in order to use their retention times (Rts) to create a calibration data file for further adjustment of Rts in samples. The accuracy improvement was based on the determination of the Kovats retention index (RI). A well-established RI value is contained in Fiehn RTL library and this is compared to the experimental RI value in order to assign a match score between the experimental and the theoretical spectra[26]. Peak detection and deconvolution were performed by retention index comparison of spectra with the Fiehn RTL library in AMDIS software v.2.69[27]. Target compounds were assessed by comparing their mass fragmentation patterns with those available in the NIST mass spectral library and Fiehn RTL library. A private library was created with the targets obtained and confirmed from retention index search correcting retention time to the experimental observed. Metabolites that were not included in Fiehn RTL library such as cholesterol and arachidonic acid were added from the NIST library. Deconvolution was optimized by a second analysis performed by the use of retention time algorithm in AMDIS. Compounds present in at least 70% of all samples (QC samples + study samples) were kept for statistical analysis.

Validation methodology

The method was validated to assess selectivity, linearity, accuracy, instrumental precision and method precision (using standards and samples), limit of detection (LOD) and limit of quantification (LOQ). Compounds were selected to represent a range of chromatographic retention times and biochemical classes. For LC-MS (8 metabolites): carnitine, α-linolenic acid, palmitic shpingosine, stearic acid, vitamin E, cholesterol, cholesteryl linoleate (CE 18:2) and triglyceride C18:2 (TG 18:2) were chosen whereas for GC-MS (10 metabolites): citrate, glucose, glutamate, glycerol, lactate, myoinositol, proline, serine, stearic acid and urea were selected. Stearic acid was included as a common standard for both techniques, as a further check of the performance of the extraction method.

The linearity of the method was assessed by assaying five different levels of concentrations of standard solutions in triplicate covering the ranges of 25 to 200% of mean values for each metabolite. Recovery was estimated by comparing in triplicate, the values of spiked samples prepared in a linear range. Within-day instrumental precision was evaluated by consecutive injections (n= 10) of a homogeneous standard solution. Intra-assay precision was tested for standards prepared in the midrange of the calibration curve and for samples from a pool of BM (n= 7). Inter-day precision was tested by repeating the experiment on a different day, with freshly prepared standard solutions and new no-defrost BM pool. Recovery was tested by spiking the standards on the sample within the calibration range (3 replicates in 3 levels). LOQ was estimated calculating signal to noise ratio (SNR) in the lowest concentration of linearity (LP1).

RESULTS AND DISCUSSION

Sample treatment and extraction protocol

After testing several extraction mixtures with different solvents, proportions and conditions, extraction protocols that used methanol/ethanol: MTBE performed better than tertiary mixtures including water in terms of giving a single phase containing a wide range of chemical classes. Here we present the optimized method.

Extraction procedures proposed and tested here, were chosen based on minimum manipulation of the sample and broad and reproducible profile of metabolites in a single phase. MTBE was chosen primarily because it is the second most widely used solvent in BM analysis used after Folch method. Previous publications[23,28-30] have shown that employment of MTBE avoids drying and resuspension steps reducing sample manipulation and human error. The high content of lipids intrinsically packed in the sample typically results in the formation of two phases. The extraction procedure used in the current study was selected based on comparing the results obtained from: (1) visual analysis of Total Ion Chromatogram (TIC), (2) IS recovery (Figure 2) and (3) total number of entities (Table S-2).

Methanol:MTBE and ethanol:MTBE systems. All ethanol/methanol: MTBE mixtures produced a single miscible phase instead of the two phases commonly observed with solvent extraction. Methanol and ethanol work as co-solvents allowing miscibility of MTBE with the water in the milk. These mixtures had the advantage of forming one phase, which not only made the procedure simpler but also reduces the potential for analytical variation and avoids the partition phenomenon of compounds when two phases are formed. The chromatograms obtained in the methanol-MTBE serial showed that with the increase of MTBE proportion, the polarity of the mix solvent is diminished and therefore the amount of non-polar metabolites and signal intensity is higher (Figure 2A). Differences were observed mainly at the beginning of the TIC after comparing methanol: MTBE proportions from (50:50) to MTBE 100%.

The selection of the extraction procedure was based on the best recovery of compounds observed throughout the chromatograph, total signal and SI recovery. The EIC of vitamin E acetate was used as the IS, this is the synthetic compound of endogenous vitamin and had a partition coefficient (log P) of 10.69 (http://www.chemspider.com/). Therefore, vitamin E acetate denotes the hydrophobicity and partition power in the media, and allowed ascertaining the recovery of a known compound with a medium polarity. As shown in Figure 2B and Table S-2, the IS and total signal were used to compare the extraction methods in order to select the best solvent proportions. Figure 2B-C and Table S-2 showed that highest recovery of IS and TS compared to others systems were achieved with methanol: MTBE (50:50).

Intentionally, an additional extraction system was tested, this was the mixture of methanol: H2O: MTBE (36:21:43µL)[23] to force the formation of two phases in the resultant systems, allowing comparison of the one- and two-phase extraction procedures. Importantly, the chromatogram from the organic phase of methanol: H2O: MTBE (Figure 2C) system showed lower intensity and less amount of compounds compared to one phase systems. The formation of the two phases in the system made the compounds split themselves into the two phases.

Finally, the extraction with methanol: MTBE (50:50) was selected for global signature profile. From this point the characterization of BM profiles by the LC–MS and GC–MS was performed accordingly to the optimized method described in the flowchart in Figure 1.

Profile characterization

LC-MS. Metabolite characterization in the chromatogram was performed using the masses with a RSD less than 20% coming from 10 independent replicates of a BM pool. In addition, features with abundance less than 10^5 and 10^4 for positive and negative mode respectively were discarded reducing the dataset dimensions for further database identification (682 out of 1283 for positive and 204 out of 938 for negative mode). Hits from data bases tended to clustered within a biochemical class and are depicted in a Total Ion Chromatogram (TIC) in Figure 3. These are summarized in Table S-3. The retention time window for the major classes of lipids shown in Figure 3A, matches the order of elution described with previous lipidomics studies[23,31]. Among the group of compounds previously reported[4,6,21,32-34]: fatty acids, glycerophospholipids, sphingolipids, sterol lipids, monoglycerides (MG), diglycerides (DG), triglycerides (TG) and cholesteryl esters (CE) were observed. Fewer compounds were detected in negative mode. Prima-

ry fatty acids, and phospholipids were detected in both ionization modes, and carbohydrates (mono and disaccharides) were identified in the dead-volume in negative mode only (Figure 3B). Interestingly, Table S-3 shows some additional classes of compounds that are not described in literature, but were tentatively identified here for BM such as fatty esters, fatty aldehydes, ceramides, glycerophosphoglycerols (PG) and glycerophosphates (PA).

Extracted ion chromatograms from the most common fatty acids; myristic, palmitic, stearic, oleic, linoleic, α-linolenic and arachidonic acid in BM consistent with previous literature[7] are presented in Figure S-1. Here, as expected, these 7 fatty acids showed different retention times in the chromatographic system. Fatty acids with longer carbon chain eluted later, while molecules with higher number of double bonds were associated with shorter retention time.

GC-MS. The BM extraction was characterized using the standard "Fiehn metabolomics retention time lock (RTL)" method. The Total Ion Chromatogram (TIC) with the metabolite assignments for 60 compounds is presented in Figure 4 and Table S-4. Metabolite classes detected included; amino acids, organic acids, fatty acids, hexose and pentose sugars, tricarboxylic (TCA) intermediates, cholesterol and disaccharides. Disaccharides were easily extracted and identified in the GC-MS profile. Here, for example, lactose represents one of the highest peaks in the chromatogram. Nevertheless, during the experimental analysis it was observed that due to their high content in the sample the filament from the EI source got dirty quickly and consequently the sensitivity of the analytical signal was decreased. So for method optimization, the filament was turned off during the time that spans disaccharide elution in the chromatogram from 23 to 26 minutes. Most of the BM compounds identified were derived from maternal endogenous metabolism. However, heptadecanoic acid is synthesized by intestinal gut bacteria[35] from maternal diet and is illustrative of the early relationship between gut bacteria and human metabolism. Other research focused on oligosaccharides; such as fucose, has shown that the primary purpose of BM oligosaccharides is to provide a nutritional source for the infant gut microbiota[20].

Method validation

Although there are no guidelines about method validation for fingerprinting in the metabolomics field, a traditional validation method for bioanalysis was performed on 10 compounds for GC-MS and 8 metabolites for LC-MS. The metabolites were chosen to cover a range of biochemical classes, polarities, functional groups, molecular weights and retention times spanning the chromatograms in both techniques. A summary of the validation parameters is provided in Tables S-5 and S-6 for LC–MS and GC–MS respectively.

For linearity, standards were fitted to the linear model ($r > 0.99$) for all selected metabolites, and no bias was found for most of them excluding; palmitic sphingosine and cholesterol for LC-MS, and urea and stearic acid for GC-MS. No practical consequences were seen in the recovery yield for any metabolite despite the bias in the four aforementioned metabolites (recoveries ranged from 90.7 to 106.1%). Recoveries ranged from 80.2 to 106.1% for all the metabolites in both techniques, except for CE 18:2 and carnitine measured by LC-MS where the recovery yield was 51.5 and 63.8% respectively. Both compounds denoted the matrix effect due to the other compounds in the sample at these particular conditions. However, although recovery did not meet the validation criteria for these two compounds, they did meet the rest of validations parameters and mainly precision which is the critical one when comparing statistically two groups. As a proof of that CE 18:2 was found significant in the biological application described below.

For LC-MS, RSD for instrumental precision for the 8 chemical standards ranged from 2.8 to 7.0%. The highest value was for TG 18:2 and it was proved to be related to the temperature in the injector that should never be lower than 15 ºC degrees to avoid precipitation of lipids. RSD for intra-assay precision for standards ranged from 1.9 to 5.9% (n= 7) and RSD for the corresponding inter-assay precision ranged from 1.9 to 5.0% (n= 14). Method precision for samples had RSD values ranging from 1.4 to 5.8% on the first day and from 3.9 to 8.2% (n = 14) for both days. Therefore precision met the expected values to be validated in all cases. Finally, the theoretical LOQ for LC-MS compounds ranged from 15 ppb for palmitic sphingosine to 389 ppb for stearic acid.

In the case of the GC-MS, RSD values for instrumental precision for the 10 chemical standards ranged from 2.8 to 7.7%. For standards intra-assay precision ranged from 2.8 to 7.9% (n= 7) and inter-assay precision from 3.6 to 8.6% (n= 14). For samples intra-assay precision RSD values ranged from 3.0 to 6.8% and from 3.9 to 10.8% in different days (n= 14). Those values can be considered acceptable in all cases. Theoretical LOQ ranged from 0.4 ppm for proline to 15.439 ppm for glycerol. There was good agreement between the BM concentration of stearic acid measured by GC-MS (20.02 ± 1.73ppm) and LC-MS (21.00 ± 1.22ppm) showing reproducibility across analytical platforms.

Network analysis of breast milk composition

Having validated the single-phase extraction method for both GC-MS and LC-MS analysis of BM, the metabolic reaction network of the metabolites identified in BM was explored using the MetaboNetworks software[36]. The metabolic reaction network highlights the interconnectivity of different domains of metabolism by considering the shortest connecting paths found in KEGG that link the identified metabolites, and reflects the composite metabolic signature of BM (Figure S-2 and S-3). In the database reactions were included if they occur in the human supra-organism, i.e. due to human enzymes and/or enzymes from the most common bacteria phyla (firmicutes, bacteroidetes, α-proteobacteria, β-proteobacteria, γ-proteobacteria, δ-proteobacteria and actinobacteria). Metabolites shown in blue were identified using LC-MS, metabolites identified using GC-MS in red and metabolites identified by both LC-MS and GC-MS in magenta. The white nodes represent the intermediary metabolites in pathways connecting the identified compounds that were not detected in the current study; these may play an important role in the internal intermediary metabolism hence they are not the final products detected by GC- and/or LC-MS analysis. In general, specific metabolite classes were detected mainly by one of the two analytical platforms employed

and often multiple metabolites were detected per biological class. For example, branched-chain amino acids, metabolites related to glycolysis and amino sugars metabolism, aromatic compounds as well as TCA cycle intermediates and TCA anaplerotic metabolites were identified with GC-MS. On the other hand, Figure S-2 highlights that metabolites related to lipid metabolism were identified by LC-MS. Cholesterol and related sterols (related to bile acid metabolism) were also found mainly with LC-MS. This shows that each technique identifies mostly metabolites associated with specific pathways and highlights the advantage of analyzing BM samples using different analytical techniques in order to achieve maximal coverage of the BM metabolome. However, for certain compound classes, such as short chain fatty acids, both techniques were capable of detection and were in good agreement.

Global profiling analysis on breast milk samples composition over time

The validated single-phase extraction method with LC- and GC-MS detection was applied to characterizing compositional differences in BM obtained 1 week and 4 weeks post-partum. Clear metabolic differences between the two time points based on spectral data obtained from LC-MS analysis were easily visualized using Principal Components Analysis (PCA). The PCA scores plot corresponding to the LC-MS data demonstrated an evolving trend in the composition of the milk along the first principal component (PC1) in relation to its time of collection post-partum (Figure S-4A). However, the corresponding PCA scores plot for the GC-MS data did not show any sample clustering in any of the first three components in relation to its time of collection (Figure S-4B).

A supervised approach based on partial least squares discriminant analysis (PLS-DA) was subsequently applied to the data obtained by LC-MS (Figure S-4C) and GC-MS (Figure S-4D) using the two sample collection time points as classifiers. Samples clustered according to class in both, the GC-MS and LC-MS PLS-DA models. This post-partum evolution in the composition of the milk is concordant with published literature[35-36]. To identify the metabolites which changed in relation to time post-partum, loading plots of both PLS-DA models were explored. These models identified a number of different metabolites changing over lactation time.

The data from LC-MS analysis identified several metabolites that increased in concentration in the samples taken more than 26 days post-partum such as linoleic acid (18:2), palmitoleic acid (16:1), oleic acid (18:1), LPE (18:2), hydroxyadipic acid, and MGs, DGs and TGs. Metabolites decreasing in concentration over this period were: lyso-and phospholipids, α-tocopherol, cholesterol and CE 18:2.

The loading plots of the PLS-DA model corresponding to the GC-MS data show oleic acid (18:1), palmitoleic acid (16:1), linoleic acid (18:2) and gluconic acid to be increased in concentration in samples collected at week 4 compared to those collected within the first 7 days post-partum. Metabolites decreasing in this period were: fucose, furanose isomers, D-glucosaminic acid and cholesterol.

The ability to establish differences in the relative abundances of metabolites present in samples, in agreement with previous literature gives further weight to the validity of the method developed and outlined in this paper. For example, previous research has similarly identified a decrease in the abundance of cholesterol from BM samples taken as lactation proceeded[37], as identified by GC-MS and LC-MS in this study. Here, the total amount of cholesterol and cholesterol esters declined over a period of 2-84 days postpartum, where the greatest difference from cholesteryl esters was observed in the content of CE 18:2.

Likewise, increases in the quantity of certain fatty acids have been described in BM samples collected at different times during lactation[38], in line with our findings. One such study examined fatty acid abundances in breast milk collected from day two to day 84 post-partum. Oleic acid 18:1, was found to be increased from 36.08± 1.03 on day two to 38.67 ± 0.84 on day 84 post-partum[38]. Similar findings were reported for the linoleic acid 18:2, with the percentage weight increasing from 13.16 ± 0.80 to 17.46 ± 0.67 over the entire period studied[38]. These literature results are all reflected by our findings from a combination of GC-MS and LC-MS, see Tables S-7 and S-8.

Concluding remarks

We have shown for the first time how BM samples can be analyzed using a single phase extraction followed by the global profiling analysis of that phase using GC-MS and LC-MS, which enables simultaneous characterization of both medium polar compounds and lipids simultaneously. Our approach is an improvement over existing methods for the analysis of BM that typically use a two phase extraction as it is more efficient in terms of time, cost and simplicity, and serves two different analytical platforms. We have applied this method to the characterization of the human BM metabolome and have used pathway mapping tools to identify representation of multiple interconnected pathways. These pathways include branched chain amino acids, aromatic amino acids, TCA cycle intermediates and anaplerotic metabolites, short chain fatty acid metabolism and lipid metabolism. Unsurprisingly, the metabolic reaction network shows that metabolites from a specific pathway are mostly identified using a single technique. However, in the case of metabolic pathways that are co-regulated by gut bacteria, both GC-MS and LC-MS are needed for identification. In future, analyzing the metabolic profile of BM using different analytical techniques will be needed to uncover the full metabolome of this biofluid. Finally, applicability of these proposed analytical methodologies was corroborated after studying changes in BM composition over time.

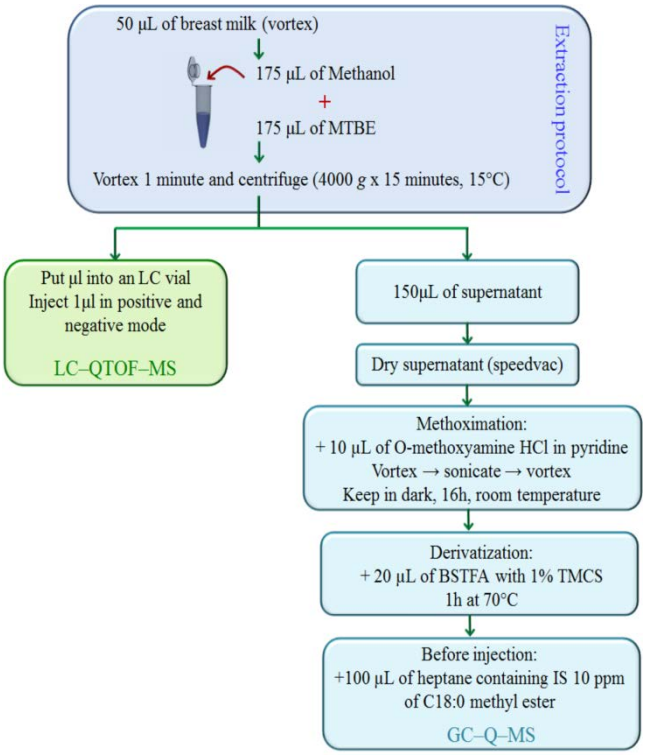

Figure 1. Optimised breast milk extraction protocol flowchart for LC–MS and GC–MS.

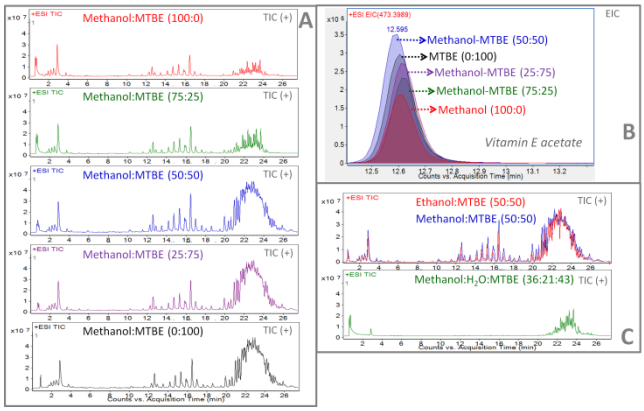

Figure 2. Graphical comparison of the different extraction mixtures tested using positive mode ionization. (A) LC-MS Total Ion Chromatograms (TIC) from methanol (100, 75, 50, 25, 0%) - MTBE (0, 25, 50, 75, 100%) proportions; (B) Extracted Ion Chromatogram (EIC) from Vitamin E acetate (*m/z*= 473.3989) for the extraction mixtures aforementioned in (A); (C) TCC comparison of methanol / ethanol – MTBE (50:50) and the TCC of the extraction mixture of methanol: H2O: MTBE (36:21:43%) showing inferior performance to the methanol: MTBE (50:50%) solvent system.

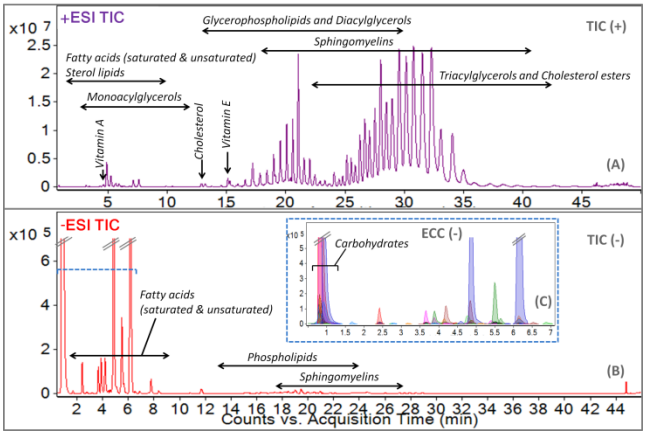

Figure 3. (A) Total Ion Chromatogram in positive mode (TIC +) and (B) TIC (-) in negative mode, acquired by LC-MS. Metabolite identification in BM profile by classes of lipid. (C) Extracted Compound Chromatogram (ECC) for negative mode showing carbohydrates. Extraction system [methanol: MTBE (50:50)].

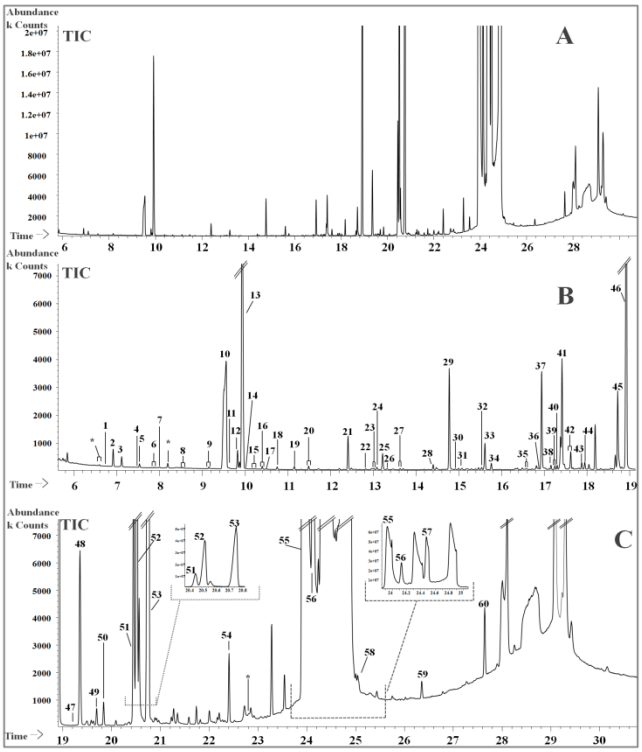

Figure 4. (A) Total Ion Chromatogram of breast milk profiling obtained by GC–Q–MS. (B) expansion of the profile from 6 to 19 minute. (C) expansion of the region from 19 min to 31 min. Lipid extraction [methanol: MTBE (50:50) proportion], for more details on analytical conditions see the text in GC-MS analysis method section, page 9. Numbers correspond to final identified compounds described in Table S-4.

AUTHOR INFORMATION

Corresponding Author

*Prof. Coral Barbas.
Centre for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy, Universidad San Pablo CEU, Campus Monteprincipe, Boadilla del Monte, 28668 Madrid, Spain.
Tel: +34 91 37 24 711
e-mail: cbarbas@ceu.es (Word Style "FA_Corresponding_Author_Footnote"). Give contact information for the author(s) to whom correspondence should be addressed.

Author Contributions

E.H., I. G-P, A.G. and C.B. designed and supervised the research A.V. and I. G-P performed milk analysis and data treatment and wrote the manuscript draft J. M. P. and M.F-L. contributed to the data analysis. I. G-P, A. J. N., and N. M. participated in the application study. A.V., I. G-P, J. M. P., E.H. and C.B. contributed to the writing of the manuscript. / All authors have given approval to the final version of the manuscript. / # Equal amount of work has been done by these authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

AV acknowledges EADS-CASA for her fellowship. Authors gratefully acknowledge the financial support from Ministry of Science and Innovation MICINN CTQ2011-23562. IGP thanks a NIHR postgraduate research fellowship. NA is supported by a doctoral research fellowship from the Chelsea and Westminster /Westminster Medical School Research Trust.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

(1) Miller, E. M.; Aiello, M. O.; Fujita, M.; Hinde, K.; Milligan, L.; Quinn, E. A. *Am J Hum Biol* **2013**, *25*, 1-11.
(2) Pons, S. M.; Bargalló, A. C.; Folgoso, C. C.; López Sabater, M. C. *Eur J Clin Nutr* **2000**, *54*, 878-882.
(3) Oozeer, R.; van Limpt, K.; Ludwig, T.; Ben Amor, K.; Martin, R.; Wind, R. D.; Boehm, G.; Knol, J. *Am J Clin Nutr* **2013**, *98*, 561S-571S.
(4) Jensen, R. G. *Lipids* **1999**, *34*, 1243-1271.
(5) Rasmussen, K. M. *Annu Rev Nutr* **1992**, *12*, 103-117.
(6) Koletzko, B.; Rodriguez-Palmero, M.; Demmelmair, H.; Fidler, N.; Jensen, R.; Sauerwald, T. *Early Hum Dev* **2001**, *65 Suppl*, S3-S18.
(7) Abrahamse, E.; Minekus, M.; van Aken, G. A.; van de Heijning, B.; Knol, J.; Bartke, N.; Oozeer, R.; van der Beek, E. M.; Ludwig, T. *Food Dig* **2012**, *3*, 63-77.
(8) Eidelman, A. I. *Breastfeed Med* **2012**, *7*, 323-324.
(9) Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X. *J Proteomics* **2012**, *75*, 1079-1088.
(10) Lenz, E. M.; Wilson, I. D. *J Proteome Res* **2007**, *6*, 443-458.
(11) Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K. *Nat Protoc* **2010**, *5*, 1005-1018.
(12) Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat Protoc* **2007**, *2*, 2692-2703.
(13) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; Consortium, H. S. M. *Nat Protoc* **2011**, *6*, 1060-1083.
(14) Marincola, F. C.; Noto, A.; Caboni, P.; Reali, A.; Barberini, L.; Lussu, M.; Murgia, F.; Santoru, M. L.; Atzori, L.; Fanos, V. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet* **2012**, *25*, 62-67.
(15) Pratico, G.; Capuani, G.; Tomassini, A.; Baldassarre, M. E.; Delfini, M.; Miccheli, A. *Natural product research* **2014**, *28*, 95-101.
(16) Giuffrida, F.; Cruz-Hernandez, C.; Flück, B.; Tavazzi, I.; Thakkar, S. K.; Destaillats, F.; Braun, M. *Lipids* **2013**, *48*, 1051-1058.
(17) Daud, A. Z.; Mohd-Esa, N.; Azlan, A.; Chan, Y. M. *Asia Pac J Clin Nutr* **2013**, *22*, 431-442.
(18) Kelishadi, R.; Hadi, B.; Iranpour, R.; Khosravi-Darani, K.; Mirmoghtadaee, P.; Farajian, S.; Poursafa, P. *J Res Med Sci* **2012**, *17*, 824-827.
(19) Zou, X.; Huang, J.; Jin, Q.; Guo, Z.; Liu, Y.; Cheong, L.; Xu, X.; Wang, X. *J Agric Food Chem* **2013**.
(20) Miller, J. B.; Bull, S.; Miller, J.; McVeagh, P. *Journal of pediatric gastroenterology and nutrition* **1994**, *19*, 371-376.
(21) Blaas, N.; Schüürmann, C.; Bartke, N.; Stahl, B.; Humpf, H. U. *J Agric Food Chem* **2011**, *59*, 6018-6024.
(22) Garcia, A.; Barbas, C. *Methods Mol Biol* **2011**, *708*, 191-204.
(23) Whiley, L.; Godzien, J.; Ruperez, F. J.; Legido-Quigley, C.; Barbas, C. *Anal Chem* **2012**, *84*, 5992-5999.
(24) Trygg, J.; Holmes, E.; Lundstedt, T. *J Proteome Res* **2007**, *6*, 469-479.
(25) Fisher, R. A. *Statistical methods for research workers.*, 7th ed.; Edinburgh Oliver and Boyd 1938, 1954, p 356.
(26) Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K.; Fraser, P. D.; Bramley, P. M. *J Exp Bot* **2005**, *56*, 219-243.
(27) Zhang, J.; Koo, I.; Wang, B.; Gao, Q. W.; Zheng, C. H.; Zhang, X. *J Chromatogr A* **2012**, *1251*, 188-193.
(28) Naz, S.; Garcia, A.; Barbas, C. *Anal Chem* **2013**, *85*, 10941-10948.
(29) Chen, S.; Hoene, M.; Li, J.; Li, Y.; Zhao, X.; Haring, H. U.; Schleicher, E. D.; Weigert, C.; Xu, G.; Lehmann, R. *J Chromatogr A* **2013**, *1298*, 9-16.
(30) Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D. *J Lipid Res* **2008**, *49*, 1137-1146.
(31) Sandra, K.; Pereira, A. o. S.; Vanhoenacker, G.; David, F.; Sandra, P. *J Chromatogr A* **2010**, *1217*, 4087-4099.
(32) Tijerina-Sáenz, A.; Innis, S. M.; Kitts, D. D. *Acta Paediatr* **2009**, *98*, 1793-1798.
(33) Canfield, L. M.; Clandinin, M. T.; Davies, D. P.; Fernandez, M. C.; Jackson, J.; Hawkes, J.; Goldman, W. J.; Pramuk, K.; Reyes, H.; Sablan, B.; Sonobe, T.; Bo, X. *Eur J Nutr* **2003**, *42*, 133-141.
(34) Crill, C. M.; Helms, R. A. *Nutr Clin Pract* **2007**, *22*, 204-213.
(35) Martysiak-Żurowska, D. *ACTA Scientiarum Polonorum Technologia Alimentaria* **2008**, *7*, 8.
(36) Posma, J. M.; Robinette, S. L.; Holmes, E.; Nicholson, J. K. *Bioinformatics* **2014**, *30*, 893-895.
(37) Bitman, J.; Wood, D. L.; Mehta, N. R.; Hamosh, P.; Hamosh, M. *Journal of pediatric gastroenterology and nutrition* **1986**, *5*, 780-786.
(38) Jackson, M. B.; Lammi-Keefe, C. J.; Jensen, R. G.; Couch, S. C.; Ferris, A. M. *Am J Clin Nutr* **1994**, *60*, 353-361.

Supporting Information (SI)

Title: Breast milk metabolome characterization in a single phase extraction, multiplatform analytical approach.

Alma Villaseñor[1], Isabel Garcia-Perez[2,3], Antonia Garcia[1], Joram M. Posma[2], Mariano Fernández-López[4], Andreas J Nicholas,[5] Neena Modi[5], Elaine Holmes,[2] Coral Barbas [1]*.

[1]Centre for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy, Universidad San Pablo CEU, Campus Monteprincipe, Boadilla del Monte, 28668 Madrid, Spain.

[2]Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ London, UK.

[3]Nutrition and Dietetic Research Group, Division of Endocrinology and Metabolism, Faculty of Medicine, Imperial College London, W12 0NN, UK.

[4]Escuela politécnica Superior. Universidad San Pablo CEU, Campus Monteprincipe, Boadilla del Monte, 28668 Madrid, Spain.

[5]Section of Neonatal Medicine, Department of Medicine, Chelsea & Westminster Hospital campus, Imperial College London, London, SW10 9NH, UK.

Table of contents:

Table S-6. Validation parameters for selected metabolites in BM by GC-Q-MS.

Table S-7. Tentative identification of significant metabolites in BM between the first week (0-7 days) against post 4 weeks postpartum by LC-QTOF-MS.

Table S-8. Tentative identification of significant metabolites in BM between the first week (0-7 days) against >4 weeks postpartum by GC-Q-MS.

Figure S-1. Extracted Ion chromatogram (EIC) in positive mode of some common fatty acids (FA) found in BM. Ions were extracted by Find by Formula algorithm (Mass Hunter-Agilent®) at 10 ppm mass accuracy from the formula and with a match score was above 80 (MFG score). Highlighted peaks were selected according to the retention time of each standard.

Figure S-2. Metabolic reaction network of metabolites found in breast milk, created using the MetaboNetworks software (Posma et al, 2014). The network shows links between metabolites if the reaction entry in KEGG shows that it is a main reactant pair and the reaction is either mediated by an enzyme linked to human genes, or in some cases bacteria, enzyme linked to genes from the most abundant endosymbionts or it is part of a spontaneous process. Metabolites shown in blue were identified using UPLC-MS, in red metabolites identified using GC-MS, in magenta identified by both LC-MS and GC-MS and in white are all metabolites linking these metabolites. The background shading illustrates different types of metabolism based on the closest affinity with some overlap between groups: tricarboxylic acid (TCA) cycle ( ); TCA anaplerotic metabolism ( ); coenzyme A metabolism ( ); lipid and fatty acid related metabolism ( ); branch-chain amino acid metabolism ( ); bile acid metabolism ( ); aromatic compounds ( ); sulphur metabolism ( ); purine and pyrimidine metabolism ( ); 1-carbon and related metabolism ( ); glycolysis and amino sugar metabolism ( ); and urea cycle metabolites ( ). An expanded version is provided in the online supplement.

Figure S-3. Expanded version of the metabolic reactions network (figure S-3).

Figure S-4. (A) PCA scores plot of the first 3 components of breast milk profiling obtained by LC-QTOF-MS comparing samples collected on days 1-7 post-partum (blue dots) vs samples collected more than 4 weeks post-partum (red dots). (B) PCA pairs plot of the first 3 components of breast milk profiling obtained by GC–Q–MS comparing samples collected on days 1-7 post-partum (blue dots) vs samples collected more than 4 weeks post-partum (red dots). (C) PLS-DA scores plot of

breast milk samples collected during the first 7 days post-partum (blue dots) vs breast milk samples collected after 4 weeks post-partum (red circles) obtained by LC-QTOF-MS (D) PLS-DA scores plot of breast milk samples collected during the first 7 days post-partum (blue dots) vs breast milk samples collected after 4 weeks post-partum (red circles) obtained by GC–Q–MS.

Reagents.

LC-QTOF-MS. Analytical grade ammonia hydroxide (30% ammonium in high purity water) was acquired from Panreac Quimica SA (Barcelona, Spain) and analytical grade methyl-*tert*-butyl-ether (MTBE) from Sigma- Aldrich (Steinheim, Germany). LC-MS grade methanol, acetonitrile and analytical grade formic acid were purchased from Fluka Analytical (Sigma- Aldrich, Steinheim, Germany). Ultrapure water was obtained from a Milli-Qplus185 system (Millipore, Billerica, MA, USA). Vitamin-E-acetate (DL-α-Tocopheryl acetate) with a monoisotopic mass of 472.3931 ($C_{31}H_{52}O_3$), was purchased from Fluka Biochemica (Switzerland) and was used as internal standard (IS).

GC-Q-MS. Standards and reagents were all of analytical grade except where stated otherwise. C18:0 methyl ester (IS), *O*-Methylhydroxylamine hydrochloride, pyridine (silylation grade), heptane, acetone, isopropanol, methanol, and acetonitrile (last three LC-MS grade) were obtained from Sigma-Aldrich (Steinheim, Germany). N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) + 1% trimethylchlorosilane (TMCS) were purchased from Pierce Chemical Co (Rockford, IL, USA). Stock reference mix solution (Grain FAME mix) was purchased from Supelco/ Sigma-Aldrich (Bellefonte, USA). Standard solution was composed of a mix of 19 fatty acid methyl esters prepared in dichloromethane [10mg/ml]. This solution was kept at -20 °C and was diluted 1/100 before GC-MS analysis.

Table S-1. List of compounds chosen for validation for LC-QTOF-MS and GC-Q-MS.

No.	Compound	Biochemical class	Log P*	RT (min)	LC-QTOF-MS	GC-Q-MS
1	**Carnitine**	Alkylamines	-4.52±0.73	0.88	+	-
2	**α-Linolenic acid**	Fatty Acids	6.50±0.33	4.66	+	-
3	**Stearic acid**	Fatty Acids	8.22±0.19	9.20 (LC) 20.63 (GC)	+	+
4	**Cholesterol**	Steroids and derivatives	9.85±0.28	13.48	+	-
5	**Vitamin E**	Prenol Lipids	11.90±0.27	15.85	+	-
6	**Palmitic Sphingosine**	Sphingolipids	13.34±0.57	17.75	+	-
7	**CE 18:2**	Steroids and derivatives	18.17	29.60	+	-
8	**TG 18:2**	Triradylglycerols	22.16±0.44	30.07	+	-
9	**Lactic acid**	Hydroxy Acids	-0.70±0.27	6.78	-	+
10	**Urea**	Ureas	-2.11±0.19	9.38	-	+
11	**Myoinositol**	Cyclic Alcohols	-2.11±0.49	19.25	-	+
12	**Proline**	Amino Acids	-0.57±0.28	10.17	-	+
13	**Glutamic**	Amino Acids	-1.67±0.47	14.29	-	+
14	**Serine**	Amino Acids	-1.58±0.33	9.93	-	+
15	**Glucose**	Monosaccharides	-3.17±0.86	17.37	-	+
16	**Citric acid**	Carboxylic Acids	-1.72±0.40	16.52	-	+
17	**Glycerol**	Sugar Alcohols	-2.32±0.49	9.83	-	+

+/- denotes whether the compound was validated (+) or not (-) by each technique. * Values obtained from: chemspider website (http://www.chemspider.com/).

Table S-2. Comparison of total number of entities between samples prepared with the mixture methanol: MTBE and ethanol: MTBE.

System	Total Signal	System	Total Signal
Methanol 100%	2.1×10^8	**Ethanol 100%**	4.8×10^8
Methanol-MTBE (75:25)	2.9×10^8	**Ethanol-MTBE (75:25)**	6.1×10^8
Methanol-MTBE (50:50)	7.1×10^8	**Ethanol-MTBE (50:50)**	6.1×10^8
Methanol-MTBE (25:75)	6.5×10^8	**Ethanol-MTBE (25:75)**	6.5×10^8
MTBE 100%	6.5×10^8		

Table S-3. Tentative characterization of compounds present in breast milk profile by LC-QTOF-MS in positive and negative mode.

Category	Subcategory	Positive Mode LC-MS	Negative Mode LC-MS	Shared
Fatty Acyls	**Fatty Acids & Conjugates**	22	33	6
	Fatty alcohols	10	1	0
	Fatty aldehydes	11	4	1
	Fatty amides	3	0	0
	Fatty esters	8	16	5
	Octadecanoids	2	1	0
	Eicosanoids	0	3	0
	Other Fatty Acyls	0	1	0
Glycerolipids	**Monoacylglycerols**	41	0	0
	Diacylglycerols	48	0	0
	Triacylglycerols	22	0	0
	Ladderans	9	0	0
Glycerophospholipids	**Glycerophosphates**	8	5	0
	Glycerophosphocholines	22	5	3
	Glycerophosphoethanolamines	24	9	6
	Glycerophosphoglycerols	2	3	0
	Glycerophosphoinositols	1	1	0
	Glycerophosphoserines	1	1	0
	Glycosylglycerophospholipids	0	1	0
Sphingolipids	**Phosphosphingolipids**	23	0	0

	Ceramides	1	0	0
	Phosphonosphingolipids	1	0	0
	Glycosphingolipids	2	0	0
	Sphingoid bases	0	1	0
	Neutral glycolsphingolipids	0	1	0
Prenol Lipids	**Quinones &hydroquinones**	2	0	0
	Isoprenoids	3	1	0
Sterol Lipids	**Bile acids and derivatives**	5	0	0
	Secosteroids	10	4	0
	Steroids	4	8	2
	Sterols	2	0	0
Carbohydrates	**Monosaccharides**	0	2	0
	Oligosaccharides	0	3	0
	Sugar Acids	0	1	0
	Sugar Alcohols	0	1	0
	Trisaccharides	0	2	0
Others	**Tripeptide**	0	13	0
	Modified dipeptide		2	
	Cofactors	0	1	0
	Organic acids	0	1	0
	Carnitines	1	0	0
	Amino Acids	1	1	0

Tentative identification was made via our website http://biolab.uspceu.com/mediator/public which search the neutral mass into three databases (METLIN, Lipid maps and KEGG). Hits pertaining to secondary plant metabolites, pharmaceutical or synthetic compounds were discarded.

Table S-4. Compounds identified in breast milk by GC-Q-MS.

No.	Compound	Target Ion (Da)	Retention time (min)	No.	Compound	Target Ion	Retention time (min)
1	Pyruvic acid	174	6.76	**31**	Lyxose 2 / Lyxosylamine 2/ Ribose	103	15.06
2	Lactic acid (Standard confirmed)	147	6.92	**32**	Xylitol	217	15.47
3	Glycolic acid	147	7.13	**33**	Fucose 1	117	15.59
4	Valine 1	72	7.43	**34**	Fucose 2	117	15.73
5	Alanine 1	116	7.53	**35**	Citric acid (Standard confirmed)	273	16.57
6	2-Hydroxybutyric acid	147	7.87	**36**	Hippuric acid 2	105	16.88
7	2-Furoic acid	125	7.99	**37**	Myristic acid	117	16.91
8	Isoleucine 1	86	8.57	**38**	Tagatose 1 / Sorbose 2 / Sorbose 1 / Fructose 1	103	17.12
9	Valine 2	144	9.14	**39**	Tagatose 2 / Fructose 2 / Fructose 1	103	17.24
10	Urea (Standard confirmed)	147	9.54	**40**	Galactose 1 / Mannose 1 / Allose 1 / Gluconic acid lactone 1	205	17.28
11	Benzoic acid	179	9.62	**41**	Glucose 1 / Talose 1 (Standard confirmed)	319	17.40
12	Caprylic acid	201	9.81	**42**	Altrose 2 / Mannose 2 / Glucose 2 / Allose 2 / Talose 2	319	17.54
13	Glycerol (Standard confirmed)	147	9.92	**43**	Mannitol / Sorbitol	319	17.87
14	Phosphoric acid	299	10.00	**44**	1-Hexadecanol	299	17.95
15	Proline 2 (Standard confirmed)	142	10.28	**45**	Palmitoleic acid	311	18.68
16	Glycine	174	10.39	**46**	Palmitic acid	117	18.88
17	Succinic acid	147	10.48	**47**	N-acetyl-D-mannosamine 1 / N-acetyl-D-mannosamine 2	319	19.19
18	Glyceric acid	189	10.74	**48**	Myo-inositol (Standard confirmed)	318	19.32
19	Serine 2 (Standard confirmed)	204	11.14	**49**	Methyl Stearate (Internal standard)	74	19.66
20	Threonine 2	218	11.48	**50**	Heptadecanoic acid	327	19.81
21	Capric acid	229	12.39	**51**	Linoleic acid	75	20.42
22	Malic acid	233	12.80	**52**	Oleic acid	339	20.48
23	Adipic acid	111	13.00	**53**	Stearic acid (Standard confirmed)	341	20.69
24	Threitol	217	13.06	**54**	Arachidic acid	369	22.37
25	Pyroglutamic acid	156	13.21	**55**	Lactulose 1	204	23.92
26	Glutamic acid 1	174	13.33	**56**	Sucrose	361	24.10
27	Creatinine	115	13.62	**57**	Lactose 1 / Cellobiose 1	361	24.46
28	Glutamic acid 2 (Standard confirmed)	246	14.37	**58**	Trehalose / Maltose 1 / Maltose 2	361	24.91
29	Lauric acid	257	14.75	**59**	Galactinol 1	204	26.30
30	Lyxose 1 / Lyxosylamine 1	103	14.89	**60**	Cholesterol	129	27.64

Compounds were assigned using the Fiehn RTL and NIST08 libraries or by addition of the corresponding standard. For conditions see text. Numbers 1or 2 refer to the number of trimethylsilyl groups in the molecule after derivatization.

Table S-5. Validation parameters for selected metabolites in BM by LC-QTOF-MS.

Validation criteria	Carnitine	Stearic acid	Vitamin E	Palmitic Sphingosine	CE 18:2	Cholesterol	α-Linolenic acid	TG 18:2
Linearity								
Slope	6.17E+05 ± 2.63E+04	4.98E+03 ± 4.14E+02	1.49E+05 ± 5.39E+03	1.27E+06 ± 8.19E+04	2.79E+05 ± 2.83E+04	4.04E+04 ± 3.97E+03	8.92E+03 ± 7.22E+02	1.25E+06 ± 9.03E+04
Intercept	(-)1.36E+04 ± 3.31E+04	(-)6.55E+02 ± 9.77E+03	(-)8.05E+03 ± 9.33E+03	5.78E+03 ± 3.87E+03	(-)1.38E+04 ± 1.72E+04	3.12E+05 ± 1.49E+05	9.79E+03 ± 2.15E+04	4.42E+05 ± 9.17E+05
R	0.998	0.991	0.998	0.995	0.992	0.990	0.993	0.995
Range (ppm)	0.24 - 1.93	4.79 - 38.33	0.35 - 2.82	0.015 - 0.07	0.15 - 1.21	9.95 - 59.68	5.89 - 47.13	1.95 - 15.60
Recovery								
(%)	63.8	102.5	89.4	98.2	51.5	99.0	81.6	82.0
RSD (%)	11.0	5.3	8.1	6.3	15.0	6.8	7.98	4.68
Method precision with standard								
Instrumental precision (n=10), % RSD	3.6	4.4	5.4	2.8	4.8	2.9	3.5	7.0
Intra-day (n=7), % RSD	1.9	5.9	1.4	2.5	4.3	2.3	3.4	2.1
Inter-day (n=14), % RSD	1.9	5.0	2.1	2.8	3.3	3.9	3.6	2.3
Method precision with sample								
Intra-day (n=7), % RSD	2.6	5.8	3.0	3.2	2.6	1.4	4.5	4.1
Inter-day (n=14), % RSD	4.3	5.8	5.3	3.9	5.7	6.6	5.4	8.2
LOD (ppm)	0.016	1.168	0.065	0.004	0.016	0.225	0.743	0.006
LOQ (ppm)	0.053	3.895	0.217	0.015	0.053	0.749	2.475	0.023
Concentration in the BM pool sample [ppm ± (SD)]								
	0.86 ± (0.04)	21.00 ± (1.22)	1.49 ± (0.08)	1.84 ± (0.07)	0.42 ± (0.02)	1.24 ± (0.08)	0.88 ± (0.05)	1.28 ± (0.11)

Table S-6. Validation parameters for selected metabolites in BM by GC-Q-MS.

Validation criteria	Lactic acid	Urea	Glycerol	Myoinositol	Stearic acid	Proline	Glutamic	Glucose	Citric acid	Serine
Linearity										
Slope	9.17E-02 ± 5.83E-03	4.03E-02 ± 3.77E-03	1.62E-02 ± 1.58E-03	5.85E-02 ± 3.90E-03	4.37E-02 ± 2.90E-03	9.02E-02 ± 7.67E-03	2.19E-02 ± 1.66E-03	3.10E-02 ± 2.05E-03	2.25E-02 ± 1.51E-03	3.64E-02 ± 2.25E-03
Intercept	1.27E-01 ± 7.15E-02	(-)1.36E+00 ±7.24E-01	(-)3.59E-01 ± 4.94E-01	(-)2.82E-01 ± 6.19E-01	3.25E-01 ± 9.67E-02	(-)2.94E-02 ± 3.17E-02	(-)1.47E-02 ± 5.33E-02	(-)1.30E-01 ± 1.97E-01	(-) 1.66E-02 ± 2.51E-02	1.68E-02 ± 1.51E-02
R	0.995	0.991	0.992	0.994	0.994	0.990	0.992	0.995	0.994	0.997
Recovery										
Range (ppm)	2.89 - 23.09	39.03 - 312.20	70.40 - 563.20	32.32 - 258.56	6.55 - 52.42	0.84 - 6.72	6.54 - 52.32	18.7 - 149.64	2.89 - 23.09	1.88 - 15.04
Recovery (%)	97.7	106.1	98.3	95.7	90.7	82.5	96.5	90.5	80.2	104.1
RSD (%)	1.8	5.1	13.1	3.0	4.0	8.0	8.0	8.0	6.9	18.6
Method precision with standard										
Instrumental precision (n=10), % RSD	2.8	3.6	3.0	3.3	5.5	7.7	5.8	3.9	3.3	6.9
Intra-day (n=7), % RSD	3.5	5.8	5.1	3.2	2.8	7.9	3.3	3.7	3.5	6.3
Inter-day (n=14), % RSD	4.5	5.3	7.8	3.6	3.4	8.6	6.9	4.1	3.6	7.9
Method precision with sample										
Intra-day (n=7), % RSD	3.5	4.1	4.0	3.8	6.5	6.8	3.0	4.7	5.8	3.9
Inter-day (n=14), % RSD	5.0	4.7	5.2	3.9	8.6	10.8	6.6	6.7	7.8	4.7
LOD (ppm)	0.562	3.216	4.632	1.93	0.394	0.133	1.510	0.531	0.204	0.696
LOQ (ppm)	1.874	10.721	15.439	6.44	1.313	0.400	5.031	1.594	0.612	2.321
Concentration in the BM pool sample [ppm ± (SD)]										
	11.96 ± (0.59)	179.68 ± (8.40)	446.66 ± (23.09)	146.44 ± (5.77)	20.02 ± (1.73)	2.49 ± (0.27)	99.58 ± (8.36)	106.78 ± (7.13)	72.92 ± (5.72)	4.73 ± (0.22)

Table S-7. Tentative identification of significant metabolites in BM between the first week (0-7 days) against post 4 weeks postpartum by LC-QTOF-MS.

1st week vs 4 weeks											
↑4 weeks						↓4 weeks					
Compound	**Neutral mass (Da)**	**Adduct**	**m/z**	**p value**	**Change (%)**	**Compound**	**Neutral mass (Da)**	**Adduct**	**m/z**	**p value**	**Change (%)**
LPE(18:2)	**477.2841**	M+H M+Na	478.2911 500.2690	0.038	58.51	Hydroxyadipic acid	**162.0528**	M+H 2M+H+1 2M+NH_4	163.0601 326.1221 342.1399	0.017	-64.98
Octadecatrienal / Ladderane-hexanol	**262.2284**	M+H+[-H_2O] M+H	245.2251 263.2361	0.045	52.15	α-Tocopherol*	**430.3789**	M+H	431.3860	0.033	-157.91
Oleic acid C18:1*	**282.2551**	M+Na	305.2453	0.018	124.96	Cholesterol*	**386.3541**	M+H+[-H_2O]	369.3508	0.044	-24.36
Palmitoleic acid C16:1*	**254.2216**	M+Na	277.2149	0.011	72.59	CE 18:2*	**648.5830**	M+K	687.5518	0.034	-69.55
Linoleic acid C18:2*	**280.2391**	M+Na	303.2292	0.011	110.98	LPC(16:0) / LPE(19:0)	**495.3312**	M+H	496.3385	0.048	-80.02
Anandamide (20:4)	**347.2815**	M+H	348.2883	0.024	57.96	LPC(15:0) / LPE(18:0)	**481.3154**	M+H M+Na	482.3227 504.3073	0.045	-90.94
Sphingosine / Ketosphinganine / Palmitoyl Ethanolamide/ Amino-octadecanoic acid	**299.2815**	M+H	300.2892	0.031	38.21	PE(40:1) / PC(40:1)	**731.5438**	M+H M+Na	732.5518 754.5307	0.044	-75.63
Pregnanetriol	**336.2652**	M+H	337.2725	0.033	59.13	PC(40:0) / PE(40:0)	**733.5608**	M+H M+Na	734.5678 756.5515	0.015	-110.70
SM(d41:1) / N-Stearoyl-D-sphingomyelin / PE-Cer(d41:1)	**730.5966**	M+H M+Na	731.6044 753.5848	0.040	42.72	PC(42:1) / PE(42:1)	**759.5763**	M+H M+Na M+K	760.5832 782.5674 798.5406	0.044	-71.68
Keto-tridecanoic acid / Methoxy-dodecenoic acid	**228.1717**	M+H	229.179	0.047	69.01	PC(42:0) / PE(42:0)	**761.5920**	M+H M+Na	762.5995 784.5841	0.040	-63.61
Keto-stearic acid / Hydroxy-octadecenoic acid	**298.2500**	M+H	299.2573	0.045	50.66	PC(44:1) / PE(44:1)	**787.6067**	M+H M+NH_4	788.6140 805.6329	0.034	-86.06
Oxo-nonadecanoic acid	**312.2657**	M+H	313.2730	0.011	76.52	SM(d39:1) /	**702.5654**	M+H	703.5725	0.023	-60.26

						PE-Cer(d39:1)		M+Na	725.5512		
Hydroxy-eicosenoic acid / Keto-eicosanoic acid	**326.2811**	M+H	327.2875	0.042	46.63	GlcCer(d48:1) / Galbeta-Cer(d48:1)	**811.6870**	M+H+[-H_2O] M+H M+Na	794.6835 812.6915 834.6765	0.029	-65.77
MG(14:0)	**302.2453**	M+H+[-H_2O] M+H M+Na M+K 2M+Na	285.2417 303.2519 325.2347 341.2085 627.4788	0.046	42.84	TG(61:2)	**942.8581**	M+NH_4 M+Na M+K	960.8922 965.8473 981.8207	0.015	-280.21
MG(15:0) / Dihydroxy stearic acid	**316.2603**	M+H+[-H_2O] M+H M+Na	299.2572 317.2672 339.2492	0.044	51.11						
MG(16:0)	**330.2772**	M+H+[-H_2O] M+H M+Na M+K 2M+Na	313.2739 331.2844 353.2669 369.2392 683.5398	0.038	44.98						
MG(18:2)	**354.2762**	M+H+[-H_2O] M+H M+NH_4 M+Na M+K	337.2724 355.2832 372.3083 377.2657 393.2392	0.040	52.27						
MG(18:1) / Heneicosanedioic acid	**356.2918**	M+H M+NH_4 M+Na M+K	357.2991 374.3247 379.2816 395.2547	0.045	53.50						
MG(18:0)	**358.3076**	M+H+[-H_2O] M+H M+NH_4 M+Na M+K	341.3041 359.3149 376.3405 381.2974 397.2704	0.041	47.58						
Hexadecyl-arachidonoyl-glycerol / Methyl-pentadecanoyl-ladderane-octanyl-sn-glycerol	**602.5253**	M+H	603.5331	0.015	56.99						
DG(31:2)	**508.4114**	M+H+[-H_2O]	491.4077	0.022	58.82						

		M+H M+NH_4 M+K	509.4165 526.4453 547.3742								
DG(31:1)	**510.4277**	M+H M+NH_4 M+Na M+K 2M+Na	511.4329 528.4612 533.4173 549.3899 1043.8412	0.035	67.39						
DG(33:1)	**538.4600**	M+H+[-H_2O] M+H M+NH_4 M+Na 2M+Na	521.4565 539.4653 556.4940 561.4495 1099.9027	0.040	50.19						
DG(35:2)	**564.4755**	M+H M+NH_4 M+Na	565.4807 582.5097 587.4652	0.029	66.25						
DG(35:1)	**566.4916**	M+H M+NH_4 M+Na M+K 2M+Na	567.4977 584.5264 589.4811 605.4529 1155.9687	0.038	52.68						
DG(36:3)	**576.4735**	M+NH_4 M+Na M+K	594.5071 599.4624 615.4390	0.039	59.05						
DG(36:1)	**580.5055**	M+H+[-H_2O] M+H M+NH_4 M+Na M+K	563.5019 581.5105 598.5395 603.4953 619.4689	0.034	60.24						
DG(36:0)	**582.5205**	M+H+[-H_2O] M+H	565.5175 583.5253	0.038	56.85						
DG(37:3)	**590.4905**	M+H+[-H_2O] M+H M+NH_4 M+Na	573.4864 591.4967 608.5250 613.4803	0.021	53.32						
DG(37:2)	**592.5079**	M+H M+NH4	593.5134 610.5415	0.028	54.94						

		M+Na	615.4965								
DG(37:0)	**596.5384**	M+H+[-H_2O] M+H M+NH_4 M+Na	579.5350 597.5433 614.5724 619.5274	0.027	56.06						
DG(39:3)	**618.5228**	M+H M+NH_4 M+Na	619.5293 636.5576 641.5122	0.022	52.37						
DG(39:2)	**620.5385**	M+H+[-H_2O] M+H M+NH_4 M+Na	603.5349 621.5443 638.5725 643.5278	0.037	50.68						
DG(39:1)	**622.5533**	M+H+[-H_2O] M+H M+NH_4 M+Na M+K	605.5499 623.5588 640.5870 645.5428 661.5146	0.038	44.83						
DG(43:7)	**666.5202**	M+H M+K	667.5263 705.4858	0.015	76.28						
TG(41:1)	**664.5623**	M+NH_4 M+Na M+K	682.5967 687.5518 703.5250	0.029	63.31						
TG(43:2)	**690.5786**	M+NH_4 M+Na M+K	708.6134 713.5682 729.5408	0.019	62.37						
TG(48:2)	**760.6563**	M+NH_4 M+Na M+K	778.6905 783.6462 799.6221	0.015	53.89						
TG(52:2)	**816.7197**	M+NH_4 M+Na	834.7551 839.7093	0.027	33.55						
TG(54:2)	**844.7529**	M+NH_4 M+Na M+K	862.7868 867.7409 883.7187	0.019	31.22						

Tentative identification of metabolites significantly associated with time in breast milk composition. *Confirmed by standard. For putative identification, all the compounds presented an estimated formula score >90% matched with the experimental isotopic pattern distribution on Mass Hunter software. Change (%); +/- increased/decreased in BM with the time, this was calculated as follows: (Average [4 week]-Average [1 week])/Average [1 week]*100). The *p* values were calculated applying Mann–Whitney U test and corrected by False Discovery Rate test.

LPC, LPE and MG –have been named with the number of carbons in the fatty acid chain attached to the backbone and the number of unsaturated bonds, e.g., LPC (16:0). PE, PC, DG and TG- have been named with the total number of carbons in the molecule and the number of unsaturated bonds, e.g., PC(42:0).

Table S-8. Tentative identification of significant metabolites in BM between the first week (0-7 days) against >4 weeks postpartum by GC-Q-MS.

1st week vs 4 weeks					
↑4 weeks			↓4 weeks		
Compound	**Ref. *m/z***	**Change (%)**	**Compound**	**Ref. *m/z***	**Change (%)**
Oleic acid C18:1	**339**	**105.96**	Fucose	**117**	-102.49
Palmitoleic acid C16:1	**311**	**79.58**	Furanose isomers	**103**	-85.22
Linoleic acid C18:2	**75**	**97.62**	D-glucosaminic acid	**291**	-73.85
			Cholesterol	**129**	-134.92

Change (%); +/- increased/decreased in BM with the time, this was calculated as follows: (Average [4 week]-Average [1 week])/Average [1 week]*100). Ref. *m/z* refers to the fragment ion mass used for quantification.

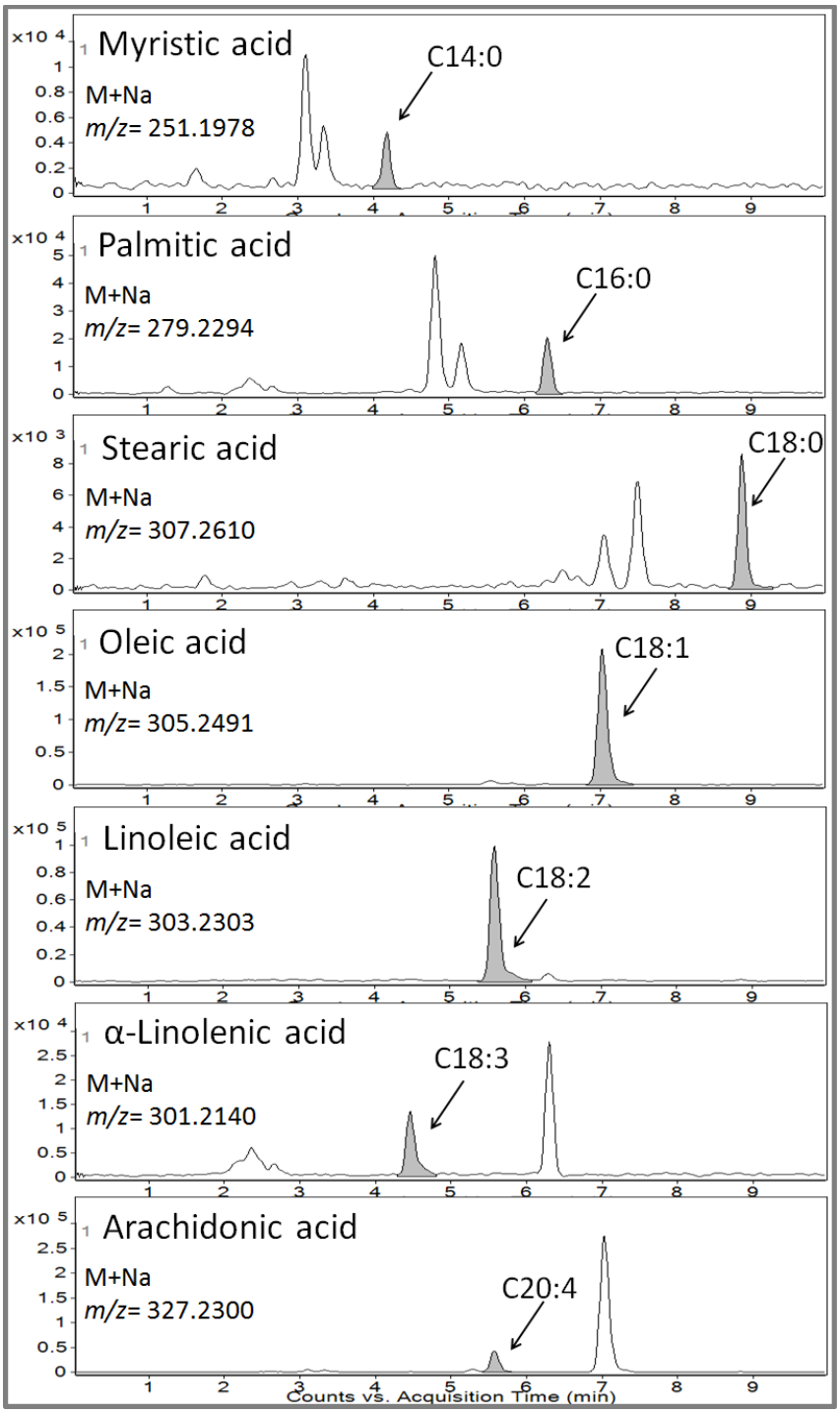

Figure S-1. Extracted Ion chromatogram (EIC) in positive mode of some common fatty acids (FA) founded in BM. Ions were extracted at 10 ppm mass accuracy from the *m/z* value. Highlighted peaks were selected according to the retention time of each standard.

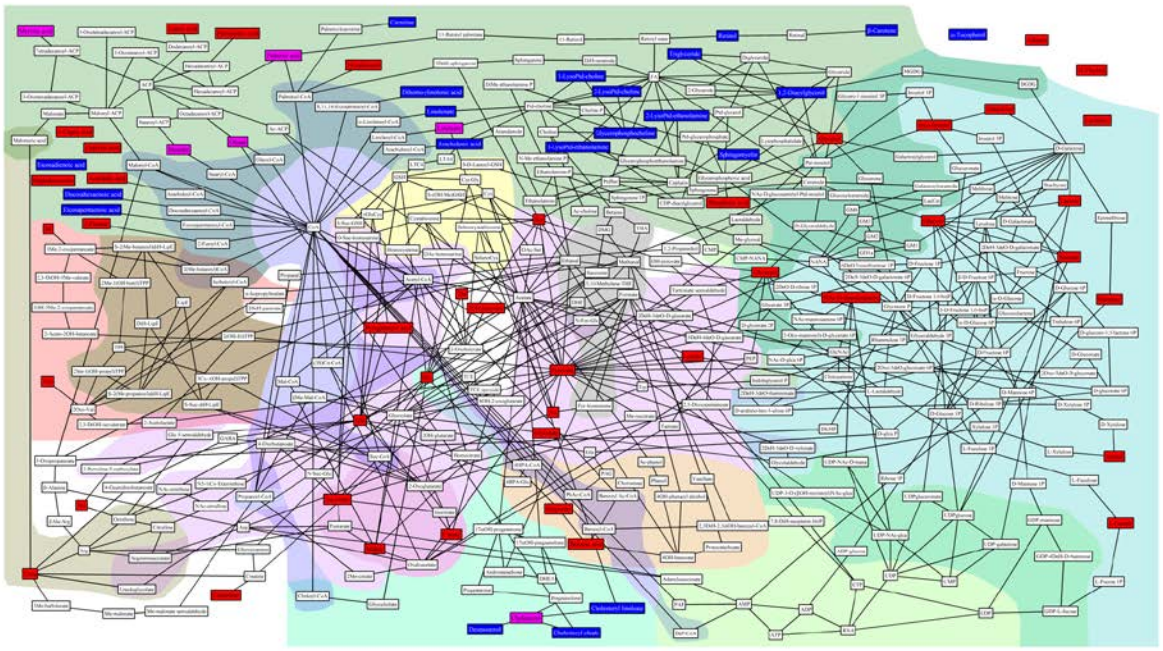

Figure S-2. Metabolic reaction network of metabolites found in breast milk, created using the MetaboNetworks software (Posma et al, 2014). The network shows links between metabolites if the reaction entry in KEGG shows that it is a main reactant pair and the reaction is either mediated by an enzyme linked to human genes, or in some cases bacteria, enzyme linked to genes from the most abundant endosymbionts or it is part of a spontaneous process. Metabolites shown in blue were identified using UPLC-MS, in red metabolites identified using GC-MS, in magenta identified by both LC-MS and GC-MS and in white are all metabolites linking these metabolites. The background shading illustrates different types of metabolism based on the closest affinity with some overlap between groups: tricarboxylic acid (TCA) cycle ( ); TCA anaplerotic metabolism ( ); coenzyme A metabolism ( ); lipid and fatty acid related metabolism ( ); branch-chain amino acid metabolism ( ); bile acid metabolism ( ); aromatic compounds ( ); sulphur metabolism ( ); purine and pyrimidine metabolism ( ); 1-carbon and related metabolism ( ); glycolysis and amino sugar metabolism ( ); and urea cycle metabolites ( ). An expanded version is provided in the online supplement.

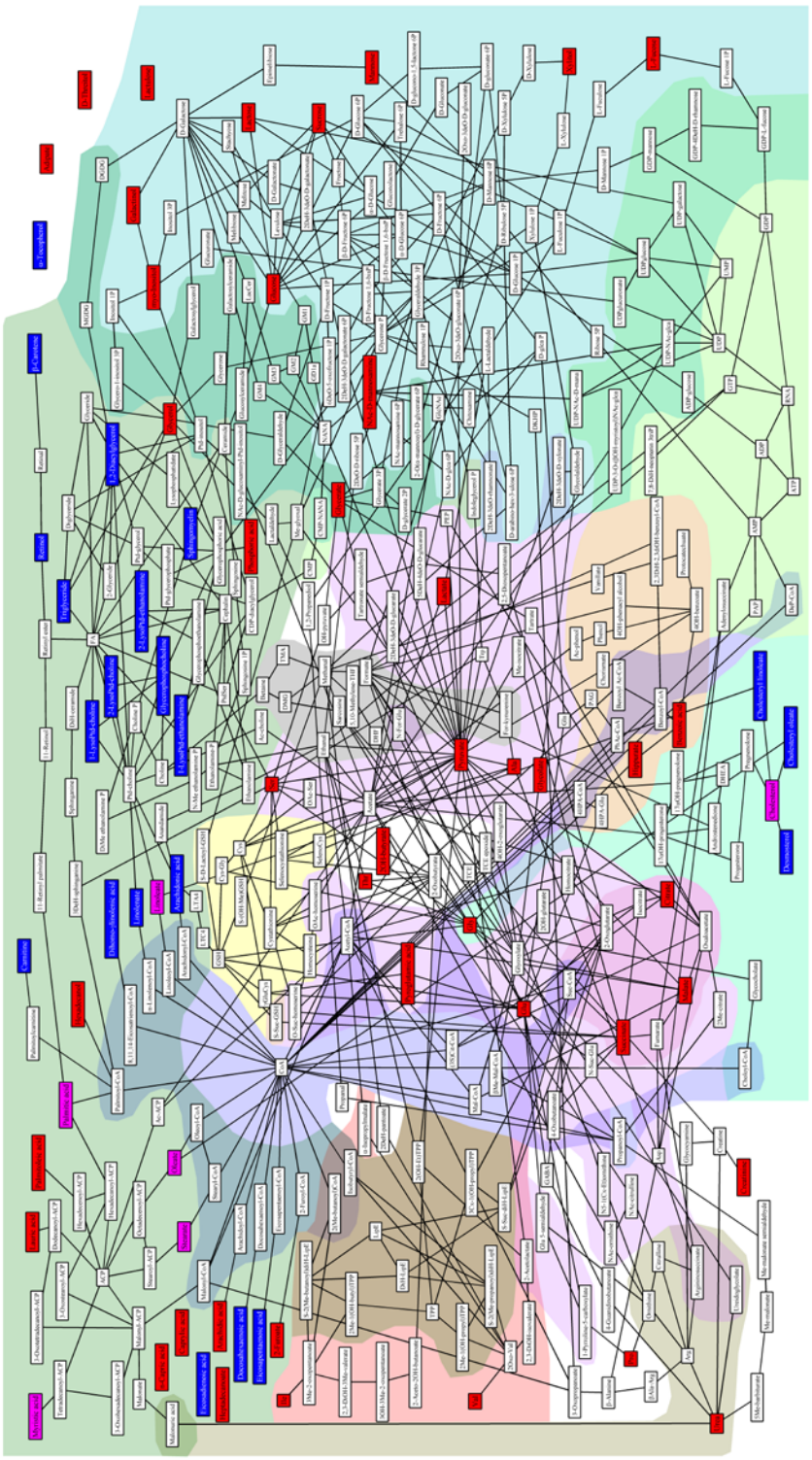

Figure S-3. Expanded version of the metabolic reactions network (Figure S-2).

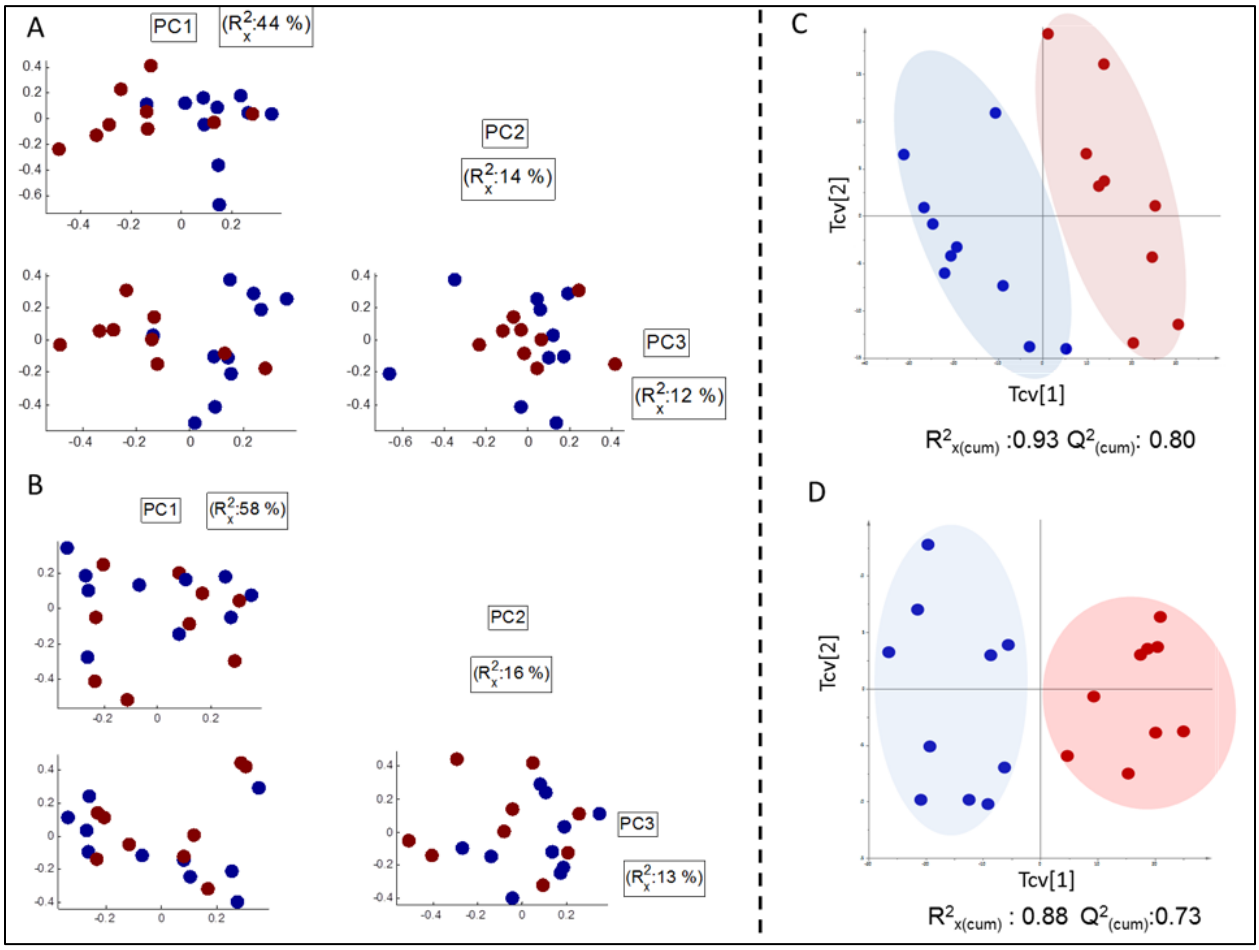

Figure S-4. (A) PCA scores plot of the first 3 components of breast milk profiling obtained by LC-QTOF-MS comparing samples collected on days 1-7 post-partum (blue dots) vs samples collected more than 4 weeks post-partum (red dots). (B) PCA pairs plot of the first 3 components of breast milk profiling obtained by GC–Q–MS comparing samples collected on days 1-7 post-partum (blue dots) vs samples collected more than 4 weeks post-partum (red dots). (C) PLS-DA scores plot of breast milk samples collected during the first 7 days post-partum (blue dots) vs breast milk samples collected after 4 weeks post-partum (red circles) obtained by LC-QTOF-MS (D) PLS-DA scores plot of breast milk samples collected during the first 7 days post-partum (blue dots) vs breast milk samples collected after 4 weeks post-partum (red circles) obtained by GC–Q–MS.