Simultaneous exploration of nutrients and pollutants in human milk and their impact on preterm infant growth: An integrative cross-platform approach

Germán Cano-Sancho, Marie Cécile Alexandre-Gouabau, Thomas Moyon, Anne-Lise Royer, Yann Guitton, Hélène Billard, Dominique D. Darmaun, Jean-Christophe Rozé, Clair-Yves Boquien, Bruno Le Bizec, et al.

Title: Simultaneous exploration of nutrients and pollutants in human milk and their impact on preterm infant growth: an integrative cross-platform approach

Authors: Cano-Sancho, German [a*]; Alexandre-Gouabau, Marie-Cécile [b]; Moyon, Thomas [b]; Royer, Anne-Lise [a]; Guitton, Yann [a]; Billard, Hélène [b]; Darmaun, Dominique [b]; Rozé, Jean-Christophe [c]; Boquien, Clair-Yves [b,d]; Le Bizec, Bruno [a]; Antignac, Jean-Philippe [a]

Affiliations:

[a] LABERCA, Oniris, INRA, 44307, Nantes, France

[b] Nantes Université, INRA, UMR1280, Physiopathologie des Adaptations Nutritionnelles, Centre de Recherche en Nutrition Humaine Ouest (CRNH-Ouest), Institut des maladies de l'appareil digestif (IMAD), F-44000 Nantes, France

[c] CHU, Centre Hospitalo-Universitaire Hôtel-Dieu, Nantes, France.

[d] EMBA, European Milk Bank Association, Milano, Italy.

*Corresponding author: Laboratoire d'Etude des Résidus et Contaminants dans les Aliments (LABERCA), UMR 1329 INRA, Route de Gachet - La Chantrerie - BP 50707 - 44307 – Nantes Cedex 3, France. Email : laberca@oniris-nantes.fr

Keywords: exposome; persistent pollutants; breast milk; preterm infants; biomarkers; nutrients; risk-benefit assessment

Abstract

Early nutritional management including fortified human breastmilk is currently recommended to fulfil the energy demands and counterbalance risks associated to preterm birth. However, little is known about the potential adverse effects of exposure to persistent organic pollutants (POPs) carried in human milk on preterm infant growth. We conducted a pilot study proving the application of an integrative analytical approach based on mass spectrometry (MS) coupled to advanced statistical models, favouring the comprehensive molecular profiling to support the identification of multiple biomarkers. We applied this workflow in the frame of a preterm infants' cohort to explore environmental determinants of growth. The combination of high resolution gas and liquid chromatography MS platforms generated a large molecular profile, including 102 pollutants and nutrients (targeted analysis) and 784 metabolites (non-targeted analysis). Data analysis consisted in a preliminary examination of associations between the signatures of POPs and the normalized growth of preterm infants, using multivariate linear regression adjusting for known confounding variables. A second analysis aimed to identify multidimensional biomarkers using a multiblock algorithm allowing the integration of multiple datasets in the growth model of preterm infants. The preliminary results did not suggest an impairment of preterm growth associated to the milk concentrations of POPs. The multiblock approach however revealed complex interrelated molecular networks of POPs, lipids, metabolites and amino acids in breastmilk associated to preterm infant growth, confirming the high potential of biomarkers exploration of this proposed workflow. Whereas the present study intended to identify simultaneously pollutant and nutrient exposure profiles associated to early preterm infant growth, this workflow may be easily adapted and applied to other matrices (e.g. serum) and research settings, favouring the functional exploration of environmental determinants of complex and multifactorial diseases.

1. Introduction

Preterm infants are at higher long-term health risks, including postnatal growth failure. Postnatal nutrition thus becomes a major strategy to counterbalance the associated deleterious effects (Agostoni et al., 2010; Boquien, 2018). Early nutritional management during the first weeks of life, including parenteral and enteral feeding with human fortified milk, has been recommended, particularly for very preterm infants, by the European Society for Paediatric Gastroenterology Hepatology and Nutrition (EPSGHAN) to improve the infant development and cognitive function (Agostoni et al., 2010). Human breastmilk is a highly complex and dynamic media, characterized by a rich composition in energy and macronutrients, but also vitamins, minerals, immunoglobulins and enzymes. Energy intake has been identified as a main independent factor associated with infant growth until the age of 2 years (Hiltunen et al., 2018), being lipids the main contributor of energy requirements of healthy infants supplying up to 45-55%. The lipid fraction of breastmilk also represents a substantial source of functional and structural lipids, such as long-chain polyunsaturated fatty acids (LC-PUFA), derived from essential fatty acids: linoleic acid (LA) and α-linolenic acid (ALA), with central roles in brain development and child growth.

The nutritional richness of breastmilk is tightly confronted by the unintended presence of xenobiotics including the lipophilic pollutants excreted within the lipid fraction, such as the vast family of persistent organic pollutants (POPs). For these pollutants, toxicologically safe levels through breastfeeding have been proposed. For instance, a tolerable daily intake levels of 1–4 pg toxic equivalents (TEQ)/kg body weight (bw) per day have been proposed for polychlorinated dibenzodioxins and dibenzofurans (PCDD/F) (WHO 2000, 2002), and a minimum risk level (MRL) is 30 ng/kg bw per day for polychlorinated biphenyls (PCBs) as group (ATSDR 2000). The last report from the World Health Organization (WHO) United Nations Environment Programme (UNEP) global monitoring survey, revealed the still pervasive occurrence of POPs in breastmilk worldwide with highest levels of PCDD/Fs and PCBs in India, but also in some European countries (van den Berg et al., 2017). In all evaluated countries, infant exposure to dioxin-like TEQs was estimated to exceed by one or two orders of magnitude the safe levels, derived from perinatal effects reported in rodents and

monkeys (van den Berg et al., 2017) . The risk-benefit analysis associated to breast milk consumption emphasized that adverse health effects of POPs in milk appear mostly subtle and transient in nature, whereas the benefits of breastfeeding are largely supported by compelling evidence associated to reduced mortality and morbidity in preterm infants with a large list of health benefits during the entire life (van den Berg et al., 2017). In any case, no formal quantitative risk-benefit analysis of breastfeeding has been performed at individual level addressing the main methodological challenges and uncertainties associated to POPs in breastmilk (Boue et al., 2018; Nauta et al., 2018).

Considering the relevance of breastfeeding in the nutritional care of preterm infants and the tight relationship between lipid/energy fraction of breastmilk and POPs, we hypothesized that POPs may be particularly associated with the different breastmilk nutrient profiles and potentially impact preterm infant growth. Hence we conducted a proof-of-concept study illustrating the application of an integrated targeted and non-targeted analytical platform for the characterization of breastmilk pollutants, nutrients and metabolites, coupled to multidimensional statistical models to identify relevant molecular signatures for preterm infant growth and their associated networks. Specifically, the sub-objectives of the pilot study were: 1) to evaluate the associations between the concentrations of POPs and nutrient profiles (i.e. nutriome) in preterm breastmilk; 2) to evaluate the associations between the concentrations of POPs in breastmilk and early growth of preterm infants; and 3) to identify integrated signatures of POPs and nutrients in human milk associated to the early growth of preterm infants accounting for the underlying correlation structures.

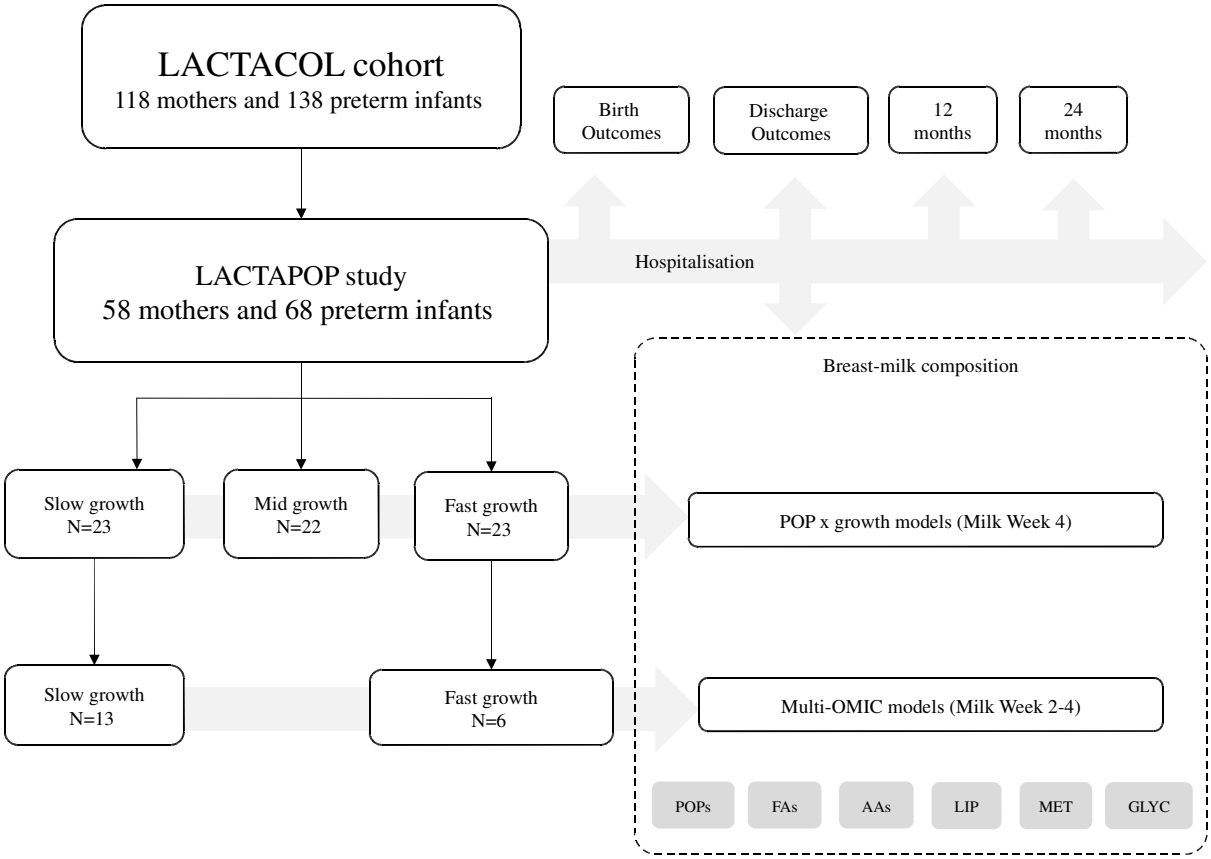
2. Material and methods

2.1. Study population

The present pilot study was conducted on a sub-group of participants selected among the mother-preterm infants' dyads enrolled in the mono-centric prospective population-based LACTACOL birth cohort (registered at www.clinicaltrails.gov as #NCT01493063) recruited from October 2011 to March 2016 in the University Hospital of Nantes, France. The milk biobank was approved by the Committee for the Protection of Persons in medical research (CPP CB-2010-03). A written consent was obtained from all participants at enrolment. One hundred thirty-eight pre-term infants with no major congenital disease except prematurity, were included, for a total of 118 mothers finally enrolled

in the LACTACOL cohort. Further details of the cohort have been published elsewhere (Alexandre-Gouabau et al., 2018). From the entire cohort, two overlapping sub-cohorts were created for this proof-of-concept study. For the first sub-cohort (LACTAPOP), a random sub-sample of 58 of mothers (delivering 68 infants, including 10 twins, born between 28 and 34 weeks of gestational age) where selected to conduct the present study involving the determination of POPs levels (Figure 1). The selection of participants was conducted without any knowledge of exposure levels and with no formal sample size calculation due to the exploratory nature of this pilot study. The 68 selected infants were ranked according to their change in weight Z-score (expressed in units of Standard Deviation (SD) and calculated as previously described (Alexandre-Gouabau et al., 2018), depicting their growth velocity during the hospital stay (i.e. difference between weight z-scores at birth and discharge) and then, were grouped in tertiles of their delta weight z-scores. For the second sub-cohort (i.e. cross-platform approach), a small group was previously selected with infants exhibiting extreme patterns of postnatal growth, belonging to the first and third tertile of weight Z-score from the entire LACTACOL cohort distribution (Alexandre-Gouabau et al., 2018). To sum up, from the preliminary selection of 26 infants (n=11, “fast growth” and n=15 “slow growth”), a subgroup of 19 infants (n=6, “fast growth” and n=13 “slow growth”) were finally included in this pilot study because complete data on POPs and nutrients in milk was available (Figure 1).

Figure 1. Study design and sampling framework. Abbreviations: POPs, persistent organic pollutants; FAs, targeted fatty acids; AAs, targeted amino acids; LIP, non-targeted lipidome; MET, non-targeted metabolome; GLYC, non-targeted glycome.



Infants admitted to the Neonatal Intensive Care Unit at Nantes University Hospital were eligible if they received human milk as their sole feeding for more than 28 days. Clinical characteristics were collected both on mothers and infants, including: maternal age, educational level, pre-gravid BMI, adverse events during pregnancy and delivery, infants' gestational age, birth weight and head circumference, growth trajectory through hospital discharge, and events during hospital stay in neonatology (Cohort characteristics in Supplemental Table 1). Volume of milk delivered per feeding session, fortifiers used and fortification level included in the parenteral nutritional supply (around 10 kcal/kg body weight/day), that generally does not exceed the first two weeks of life, as well as enteral intake, predominantly with expressed own mother's breast milk, was recorded daily. Energy and protein contents in preterm standard formula were 72 kcal/100 mL and 2 g/100 mL, respectively. Preterm infants received parenteral nutrition and minimal enteral feeding, predominantly with expressed breast milk, for a minimum of two weeks. By taking into account native own mother's milk macro-nutriments contents and routine fortification, the mean total enteral intake, particularly total energy intake, were reasonably close to that of EPSGHAN guidelines regarding daily and

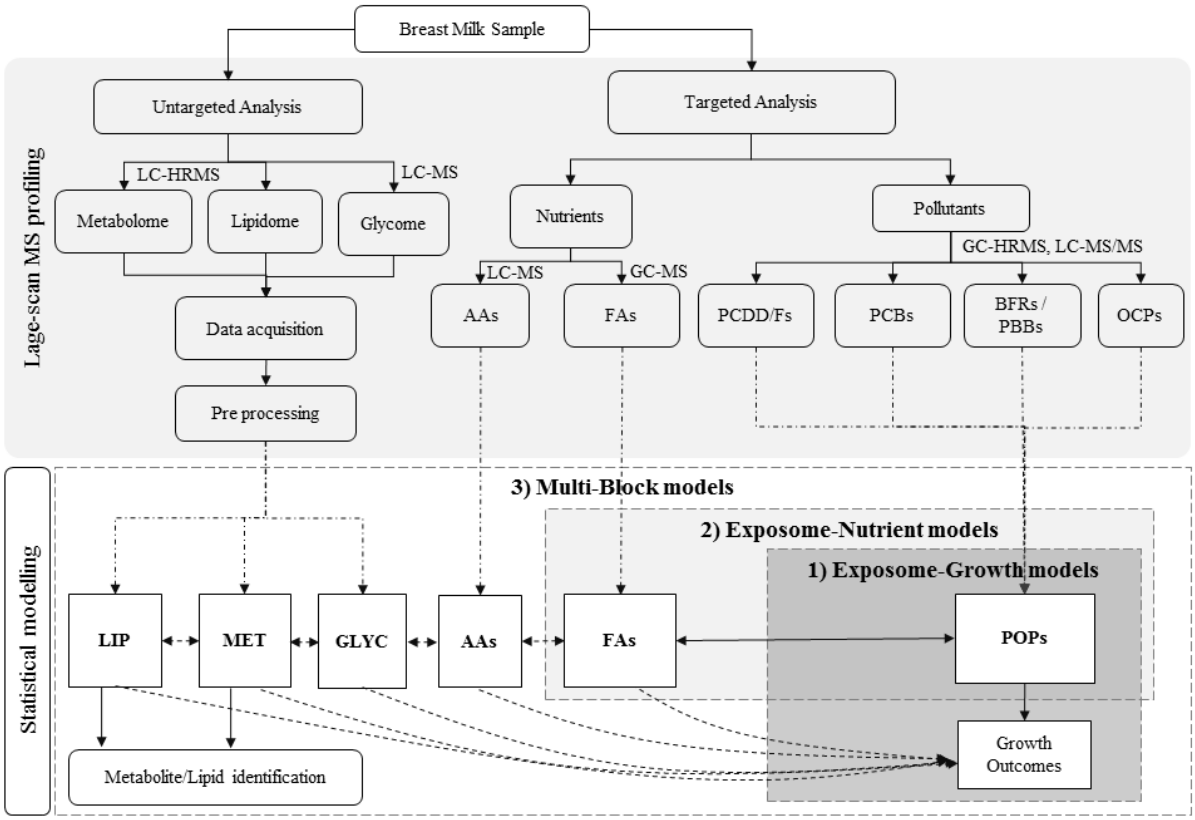
macronutrient intake, except for protein intake (which was 28% below recommended intake) (Alexandre-Gouabau et al, Nutrients 2018).The median energy intake was 124 kcal/kg body weight/day in both, the “faster growth” and the “slower growth” group, being 104 kcal/kg body weight/day in the “medium growth” group (p-value = 0.29). Anthropometrical evaluation of infants was registered at follow-up visits conducted at 12 and 24 months.

2.2. Human Milk Collection

Breast milk expression was performed manually by mothers at home, using a Medela Manual Breast pump (Medela Inc., Etampes, France). Weekly representative 24-h mature milk samples were obtained by pooling breast milk sampled from 5 to 6 bottles brought daily to the Milk Bank of the Nantes University Hospital. The whole milk pool was homogenized with a disruptor (Polytron, Lucerne, Switzerland) and kept frozen at 80°C until untargeted and targeted analysis of nutriome at week 2 and week 4 of lactation. Preliminary analysis using principal components showed no marked trajectories on nutrients during the hospital stay, supporting the use of these time points. A 30ml-breast milk expression was specifically collected at once, at discharge and frozen at −80 °C until the analysis.

2.3. Breast-milk lipidomic profiling

Figure 2. Analytical workflow for the targeted and untargeted breastmilk profiling and statistical analysis for the molecular profiling and biomarkers identification. Abbreviations: AA, amino acids; BFR, brominated flame retardants; FAs, fatty acids; GC, gas chromatography; GLYC, glycome; HRMS, high resolution mass spectrometry; LC, liquid chromatography; LIP, lipidome; MET, metabolome; MS, mass spectrometry; OCP, organochlorinated pesticides; PBBs, polybrominated biphenyls; PCBs, polychlorinated biphenyls; PCDD/Fs, polychlorinated dibenzodioxins and dibenzofurans.



Breast milk Liquid Chromatography-High-Resolution-Mass Spectrometry (LC-HRMS)–based lipidomic (LIP), metabolomic (MET) and glycomic (GLYC) untargeted profiling

Milk aliquots (50 µl) were submitted to the following preliminary treatment before analysis: 50 µl of sodium chloride 150 mmol/L were added to each milk sample followed by 100 µl of methanol and then, 100µl of chloroform, according to a modified "Bligh-Dyer" extraction. Then, two extractions with 100 µl chloroform were performed. After centrifugation (10,000 g for 10 min), the organic and aqueous layers were collected separately, dried under N2and subsequently reconstituted, in acetonitrile-isopropanol-water (ACN: IPA: H_2O 65:30:5, v/v/v), and in water-acetonitrile (H2O: ACN 95:5, v/v), for lipid (lipidome) and polar (metabolome) species, respectively. A 1200 infinity series® high performance liquid chromatography (HPLC) system (Agilent Technologies, Santa Clara, California, USA) coupled to an Exactive Orbitrap® mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) was used for both lipidomic and metabolomic profiling as previously described (Alexandre-Gouabau et al., 2019). Briefly, lipid species separation was performed on a reverse phase CSH® C_{18} (100 x 2.1 mm^2 i.d., 1.7 µm particle size) column (Waters Corporation, Milford, MA).

Sample metabolomics fingerprinting was performed, on the same LC-HRMS system, on a reverse phase with a Hypersil GOLD C18 column (1.9 µm particle size, 100x2.1 mm). The precision associated with sample preparation and LC-HRMS measurement was determined on the basis of a quality control (QC) consisting of a pool of 10 mothers' milk provided by the milk bank of Nantes Hospital Center. QCs samples were extracted similarly than others samples and injected sequentially in-between the milk samples. The tight clustering of QC samples before and after normalization were checked on principal component analysis (PCA) score plot. The extraction and reduction of human milk oligosaccharides (HMO) were performed as previously described, following milk delipidation by centrifugation and NaBH4-reduction (Alexandre-Gouabau et al., 2019; Oursel et al., 2017). Milk reduced oligosaccharides were separated on a Hypercarb® column (2.1 mm i.d. × 100 mm, 3µm particle size, Thermo Scientific, San Jose, CA, USA) on an Ultimate 3000 HPLC system coupled to an LTQ-Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA, USA).

Lipidomic, metabolomic and glycomic data analysis and metabolites characterization

As previously described (Alexandre-Gouabau et al., 2018; 2019), lipidomics and metabolomics raw data files were preprocessed with Xcalibur 2.2® (Thermo Fisher Scientific, San Jose, CA, USA), Then, lipidomics and metabolomics data were extracted using pre-processing within Workflow4Metabolomics® (W4M) (http://workflow4metabolomics.org) for automatic integration for each detected features (ions of given mass-to charge ratio and retention time) combined with CAMERA®, for annotation of isotopes and adducts and for normalization of intra- and inter-batch effects using QC samples. The resulting features [m/z; RT]) for each sample was subsequently manually sorted out according to their quality of integration and filtered by a 30% relative SD cutoff within the repeated pooled QC injections. Thereafter, accurate mass measurement of each putative metabolite was submitted to LIPID Metabolites And Pathways Strategy (LipidMaps®, www.lipidmaps.org), Human Metabolite Data Base (HMDB®, www.hmdb.ca), Biofluid Metabolites Database (MetLin®, metlin.scripps.edu) and milk metabolome database (MCDB®, www.mcdb.ca) annotation. The GlycoWorkBench software was used to draw HMOs structures. Moreover, the use of all ion fragmentation, when reverse phase chromatography was applied, helped us identify the

proposed lipids, metabolites by examination of the (pseudo) tandem mass spectrum generated, combined with the use of in-house reference databanks (Ferchaud-Roucher et al, 2015; Courant et al, 2012). Metabolite's identification level was level one, for metabolites definitively annotated with our home data base (i.e., based upon characteristic physicochemical properties of a chemical reference standard (m/z,RT) in our in-house reference databanks and their M/MS spectra compared to those of breastmilk QC) or level two, for metabolites putatively annotated (i.e., without chemical reference standards, based upon physicochemical properties and MS/MS spectral similarity with public/commercial spectral libraries, e.g.,LipidMaps®,MetLin®,and MCDB®).

2.5 Breast milk targeted free amino acids (FAAs) and total fatty acids (FAs) analysis.

FAA and FA concentrations were determined in expressed breast milk samples, collected at week 2 and week 4 of lactation. FAAs were separated and quantified, using Acquity H-Class® Ultra Perfomance Liquid Chromatography system (Waters Corporation, Milford, CO, USA) combined to a Xevo TQD® mass spectrometer (Waters Corporation, Milford, COMA, USA), following derivatization using AccQ®TagTM Ultra reagent (Waters Corporation,Milford, USA), as previously described (Alexandre-Gouabau et al, 2019). FAs concentrations were determined using the modified liquid–liquid extraction method of Bligh-Dyer and were analyzed by gas chromatography using an Agilent Technologies 7890A® instrument (Perkin Elmer, France), following trans-esterification previously described (Alexandre-Gouabau et al, 2018).

2.6. Breast-milk targeted profiling of persistent organic pollutants

The methodologies applied to isolate, detect, and quantify the targeted POPs including dioxins (17 PCDD/F congeners), polychlorobiphenyls (PCB) (12 dioxin-like and 6 non-dioxin-like PCB congeners), polybromodiphenylethers (8 PBDE congeners), polybromo biphenyls (3 PBB congeners) and organochlorine pesticides (30 OC compounds) have been previously described (Antignac et al., 2009; Costera et al., 2006; Cariou et al., 2005; Antignac et al., 2009; Bichon et al., 2015). Briefly, ^{13}C-labelled congeners were added to each sample for quantification according to the isotopic dilution method. Breast milk samples were first submitted to a liquid/liquid extraction with pentane. Resulting extracts were weighed to measure fat content (gravimetric method) and reconstituted in hexane for further sample clean up. Gel lipid compounds originated from the matrix.

For other targeted substances, three successive acid silica, florisil, and celite/carbon columns were applied for lipid removal, fractionation and further purification, respectively. PCDD/F, PCB, PBDE and OC measurements were performed by gas chromatography (Agilent 7890A) coupled to high-resolution mass spectrometry (GC-HRMS) on electromagnetic sector instruments (JEOL MS 700D or 800D®), operating at 10000 resolution and in the single ion monitoring (SIM) acquisition mode. Hexabromocyclododecane (HBCD) isomers were quantified using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) on a triple quadrupole instrument (Agilent 6410®). All these methods were fully validated according to current European criteria in the field of regular control of foodstuff of animal origin and accredited according to the ISO 17025 standard. Quality assurance/control (QA/QC) procedures included systematic analysis of negative and positive control samples in each batch of analyzed samples. Recoveries were classically calculated according to the rules defined at regulatory level (Commission Regulation 2012) and were in the 80-120% range. The method's extended uncertainty was lower than 20%. In order to investigate not only contamination levels but also contamination patterns, all determined POPs concentrations were secondly expressed as relative contributions (%) of each particular congener to the sum of all congeners from the same family.

2.7. Postnatal POP exposure estimation

The infant *i* postnatal exposure to the contaminant *n* ($E_{T2,n,i}$, ng/kg bw/day*)* from breast milk*,* was estimated by the combination of the individual normalised breastmilk intake (INT_i, mL/kg bw/day), the lipid normalised concentration in breastmilk sample *p* of contaminant *n* (*POPn*; ng/g lipids) and the specific content of lipids in breastmilk sample *p* (LIP; g lipids/100 mL milk) as follows:

$$E_{T2,i,n} = INT_i \; x \; POP_{p,n} \; x \; LIP_p$$

This approach is analogous to the estimation of the estimated daily intakes (EDI) used by WHO and allows the comparison with the safety values like the tolerable daily intakes (TDI) for given contaminants (WHO 2000). These estimates only consider the exposure from human milk, not eventual contamination of formulas, that would be expected to be similar for all infants.

2.8. Statistical analysis

Individual characteristics were summarized with median and interquartile range, for continuous variables, and frequency and percentage, for categorical variables. The tertiles of delta weight Z-score (difference between weight Z-score at birth and hospital discharge) was used to rank the infants in 3 groups: “Slower growth”, “Medium growth”; “Faster growth”. The statistical comparison of individual characteristics between groups (i.e. faster, medium and lower growth) was performed using Kruskal Wallis Test and Fisher’s exact test, for numerical and categorical variables, respectively. The distributions of POPs among groups were summarized by medians and interquartile ranges and compared statistically as continuous variables. Those chemicals with detection rates below 75% were excluded from the statistical analysis. For four analytes (PBDE 28, 154, 183 and 209) whose detection rates were between 93 and 98%, the non-detected values were replaced by limits of detection (Antweiler 2015).

The bivariate associations between milk POPs and fatty acid variables were calculated with Spearman’s correlation coefficients (ρ) and plotted them in heatmaps. Furthermore, sparse partial least squares (sPLS) was conducted to identify association patterns between milk POPs and nutrients (FAs and macronutrients) considering the correlation structure of both datasets, and displayed by means of network plots (Chun and Keles 2010). The association between the infant exposure to POPs (pg or ng/Kg BW/day, log-scale) and infant growth during hospital stay (Delta Z-score) was estimated through multivariate linear regression considering known confounding variables such as maternal age, maternal body mass index (Kg/m^2, log-scale), birth weight (grams, log-scale), gestational age (weeks), newborn sex, hospital stay duration (days) and energy intake (Kcal/kg/day, log-scale) from complementary parenteral and enteral nutrition. The two-way product interaction terms between considered variables were evaluated for statistical significance. The associations are summarized using the beta regression coefficients and 95% confidence intervals.

For a sub-group of 19 infants with duplicated breast milk sample measurements (Week 2 and Week 4, n=38) of lipidome (LIP), metabolome (MET), glycome (GLY), amino acids (AAs) and fatty acids (FAs), we extended the analytical framework towards a more comprehensive approach to integrate the chemical complexity of breastmilk in the growth models. Assuming the computational limitations associated to the minimal sample size of this sub-group, in this section we mainly aimed to set-up and

evaluate the applicability of a large-scan spectrometric platform coupled to multi-block models for a more comprehensive exploration. The results associated to the nutritional fraction has been published elsewhere, thus in this publication we will emphasize the interplay between POPs and the different nutritional fractions (Alexandre-Gouabau et al., 2018; 2019). In this regard, we finally integrated the six pre-processed (normalized and standardized) data-blocks from targeted methods, including POPs (n=52 variables), FAs (n=34 variables), AAs (n=16 variables), and the matrices from non-targeted methods, GLY (n=75 variables), LIP (n=488 variables) and MET (n=221 variables) against the infant growth (weight delta Z-score between birth and discharge) dichotomized to generate two groups of infants ("faster growth" and "slower growth"). The group of "medium" growth was not considered in this pilot study favouring the identification of molecular signatures discriminating the more extreme growth trajectories. The MS-based metabolomics and lipidomic features were previously submitted to a throughout pre-processing and filtering process of raw spectral data using Analysis of Variance-Partial Least Square (AoV-PLS) (Ghaziri et al., 2015) to reduce the internal redundancy and to pintpoint the most discriminant polar and apolar features providing a clear separation between the two infant growth groups, obtaining the final sets of 488 lipidomic and 211 metabolomic variables(Alexandre-Gouabau et al., 2018; 2019) . The datasets were integrated using the sparse generalized canonical correlation models (Tenenhaus and Tenenhaus, 2014) extended to a discriminant mode so-called DIABLO (Data Integration Analysis for Biomarker discovery using Latent variable approaches for 'Omics studies) workflow implemented with the MixOmics R package (Rohart et al., 2017; Singh et al., 2019). Briefly, similar to the partial least-squares regression, the model projects latent variables (i.e. components) based on linear combinations of underlying features maximally correlated across data-blocks and the response variable, performing variable selection via L1 penalisation (Tibshirani, 2011) to support the data-driven variable selection. The main parameters of the model are the correlation matrix design that pre-establish the degree of association known between blocks (between 0 and 1), number of components and degree of penalisation (number of variables to retain) for each component and block. The number of components is optimized through another 10-fold cross-validation (CV) process seeking at minimizing the misclassification model error

rate whereas the degree of penalisation is also driven by an internal CV procedure. The statistical analysis was conducted entirely using the R software, v. 3.5.0, with the MixOmics R package.

3. Results

3.1. Levels of contaminants in breast-milk and infant post-natal exposure estimates

The concentrations of POPs are summarized in Supplemental Table 2 expressed in lipid basis, confirming the presence of all monitored POPs in the analysed breastmilk samples. Mean dioxin toxic equivalent factors (TEF) of 6.17 and 4.50 WHO TEF_{2005}/g l.w. were observed for the sum of PCDD/Fs and dioxin-like PCBs, respectively. The mean sum concentrations of non-dioxin like PCBs, PBDEs and PBB153 resulted in respective values of 87.27, 1.7 and 0.08 ng/g l.w. Strong positive Spearman's rank bivariate correlations were noticed between POPs from the same family, whereas weaker correlations were noticed between family comparisons, especially for the case of PBDEs (Figure S1).

3.2. Levels of macronutrients, fatty acids and correlations with POPs

The summary of levels for macronutrients and FAs in breastmilk can be found in Supplemental Table 3. The correlation analysis of targeted lipophilic chemicals in breastmilk revealed strong positive associations between ω3 PUFAs (docohexaenoic acic or DHA, docosapentaenoic acid or DPA and eicosapentaenoic acid or EPA) and most POPs, most substantially for TCDD, non-coplanar and non-dioxin like PCBs (Figure 3). These toxicants were also strongly correlated to saturated FAs such as pentadecanoic acid (15:0), palmitic acid (16:0), stearic acid (18:0) and arachidic acid (20:0). Some strong inverse correlations were also noticed for ω6 PUFAs and non-coplanar PCBs, and specially docosapentaenoic acid ($22:5_{n-6}$), docosadienoic acid ($22:2_{n-6}$) and linoleic acid ($18:2_{n-6}$) exhibiting the strongest negative correlation with most POPs. The bivariate correlation analysis was extended with sPLS allowing the efficient integration of two correlated matrices and the identification of the most relevant variables and their networks, displayed in Figure 4. The relevance networks display the topological associations between FAs and POPs considering a cut-off level of bivariate correlation of 0.34 to favour the visualization of most relevant associations. This approach showed that FAs, 15:0 (pentadecanoic acid) and 20:1 (eicosenoic acid) appeared as central nodes on the association network with PBDEs, highlighting the strong negative associations between 15:0 and PBDE99 and positive associations between 20:1 and PBDE99 or the sum of PBDEs. Another network was built around total lipids and energy, negatively associated to PCBs and around the saturated FAs 20.0 (arachidic acid) and 18.0 (stearic acid) displaying positive associations with coplanar PCBs such as PCB114.

Figure 3. Heatmap representing the bivariate correlations between macronutrients, saturated (SAT), monounsaturated (MUFAs) and polyunsaturated (PUFAs) fatty acids with persistent organic pollutants in breast-milk. The Spearman correlation coefficients are represented by the colour scale. The grouped variables as sum of individual concentrations are identified by an asterisk.

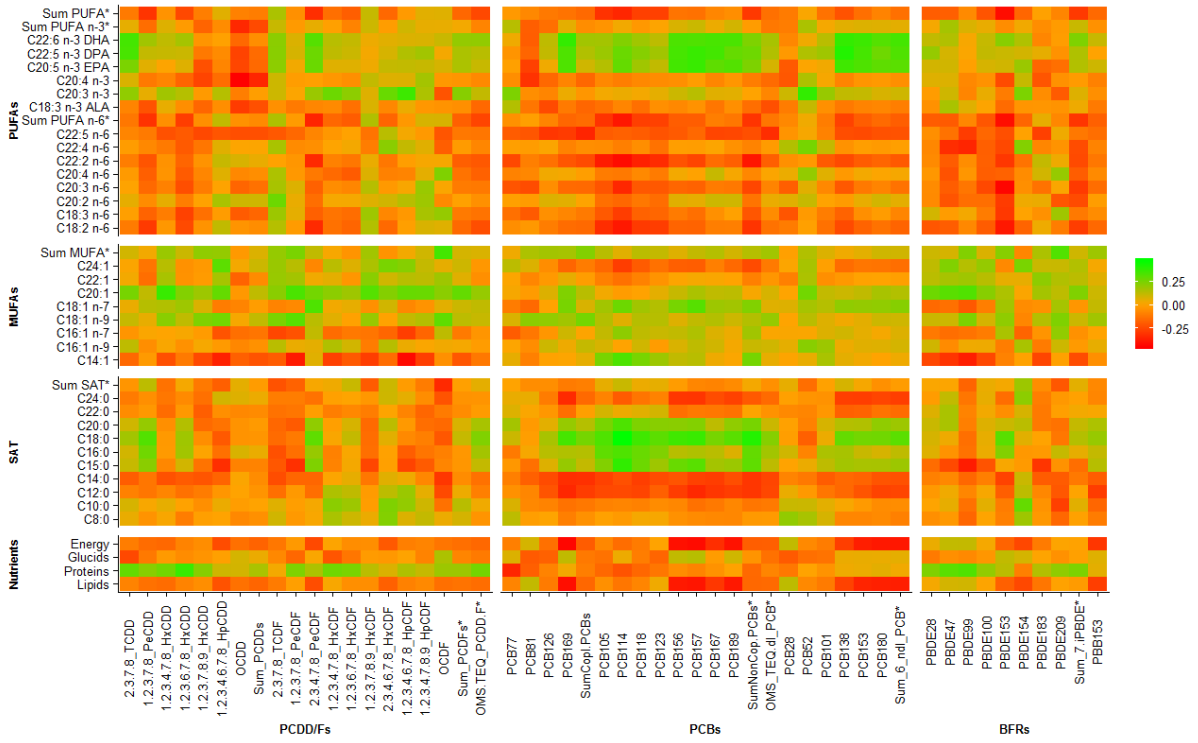
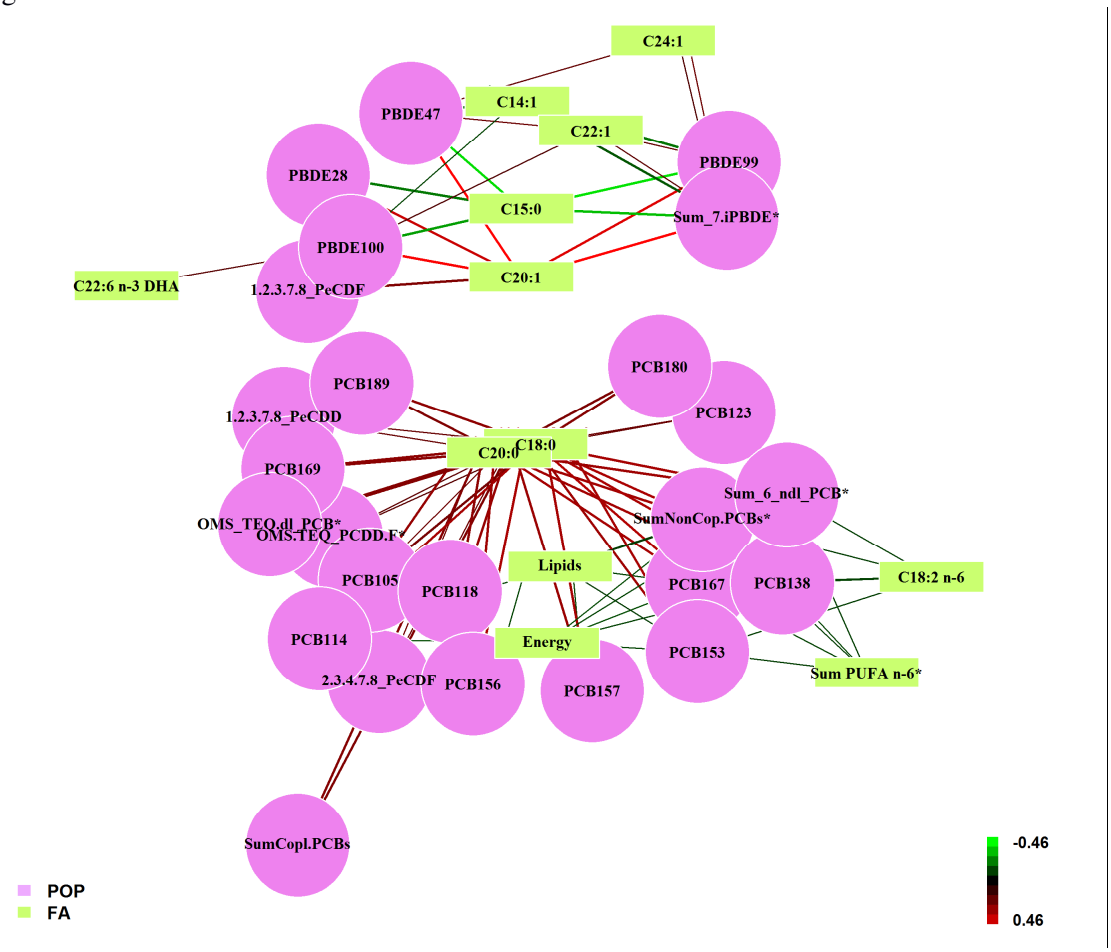


Figure 4. Relevance network plots representing the strongest associations between fatty acids (green boxes) and pollutants (purple boxes) using sparse partial least squares. The strength of the correlations between pollutants and fatty acids is represented by colour key of edges. The threshold of correlations was set up at -0.28 (negative associations, black-green) and 0.28 (positive associations, red) in order to highlight the most relevant clusters.



3.3. Associations between POPs and preterm infant growth

First, we explored the distribution of estimated daily intake of POPs through maternal breastmilk intake during the hospital stay for the infant growth groups ("Slower", "Medium" and "Faster" growth) established by the tertiles distributions of weight z-score difference between birth and hospital discharge (weight Delta z-score). The Estimated Dietary Intake (EDI) of POPs for the different Delta z-score are summarized in Table 1, showing higher concentration levels of several POPs among the "Faster" group, especially dioxin-like PCBs (e.g. PCB 126, 127, 157). In order to account for known confounding variables related to infant growth, we further conducted multivariate linear regression considering maternal age, BMI, birth weight, hospitalisation days and energy intake as covariates. The regression coefficients (and 95% confidence intervals) for the different pollutants and models are displayed with forest-plots in the Figure 5, showing the observed associations between the different POPs and infant growth. Overall, the results showed the strong confounding effect of energy intake underlying the positive association between the exposure to POPs and preterm infant growth during the hospital stay (Model 2). After the adjustment by energy intake, some chemicals exhibited positive significant associations (i.e. PCB101, 105, 118, 126, 138) with postnatal preterm infant growth. The EDI of POPs through breast-feeding was also modelled against the infant weight z-score at 24 months showing positive associations for a substantial number of dioxins after the adjustment for the confounding factors including, energy intake (Figure 6). The larger associations were depicted by 2.3.4.7.8-PeCDF ($p=0.009$) and the sum of PCDD/Fs WHO-TEQ ($p=0.01$). Some furans such as 2.3.4.6.7.8-HxCDF were also found consistently associated with preterm infant growth throughout both growth time points.

Table 1. Summary of estimated dietary intake (Median (Interquartile range) of persistent organic pollutants, macronutrients and energy from breast milk for preterm infants with slower, medium and faster growth trajectories. Infants were grouped in tertiles of weight z-scores difference between birth and discharge. P-value was estimated using Mann-Whitney-Wilcoxon Test.

	Slower N=23	Medium N=22	Faster N=23	P-value
Macronutrients and energy daily intake				
Energy enteral total [a]	123.7 (79.1 - 130.4)	104.4 (83.6 - 133.6)	124.3 (111.3 - 135.1)	0.29
Proteins enteral total [b]	2.5 (1.4 - 2.7)	1.8 (1.3 - 2.6)	2.4 (1.9 - 2.7)	0.21
Lipids enteral total [b]	5.2 (3.7 - 6.5)	5.7 (3.5 - 7.1)	5.6 (5.0 - 7.5)	0.31
Carbohydrates enteral total [b]	14.3 (8.8 - 15.6)	10.2 (7.9 - 13.6)	14.8 (12.2 - 15.4)	0.099
Energy breast-milk [a]	91.0 (72.1 - 102.9)	99.7 (77.8 - 118.6)	107.3 (101.2 - 112.1)	0.070
Proteins breast-milk [b]	1.3 (1.1 - 1.7)	1.4 (1.0 - 1.7)	1.6 (1.4 - 1.7)	0.091
Lipids breast-milk [b]	4.5 (3.7 - 5.3)	5.7 (3.3 - 7.1)	5.6 (5.0 - 7.1)	0.085
Carbohydrates breast-milk [b]	10.8 (7.1 - 11.7)	10.2 (6.5 - 11.9)	11.8 (9.7 - 12.3)	0.081
Persistent organic pollutants daily exposure				
2.3.7.8-TCDD [c]	1.8 (1.3 - 3.3)	2.0 (1.3 - 2.7)	2.7 (1.6 - 4.6)	0.072
1.2.3.7.8-PeCDD [c]	6.4 (4.2 - 11.0)	7.5 (5.4 - 9.1)	11.5 (5.8 - 14.0)	0.097
1.2.3.4.7.8-HxCDD [c]	2.8 (2.0 - 4.7)	2.8 (2.3 - 3.8)	4.2 (2.9 - 5.5)	0.060
1.2.3.6.7.8-HxCDD [c]	15.6 (10.3 - 25.2)	18.2 (13.4 - 23.9)	27.7 (14.2 - 37.8)	0.16
1.2.3.7.8.9-HxCDD [c]	3.9 (2.4 - 5.3)	3.7 (2.6 - 4.5)	5.2 (3.7 - 8.0)	0.087
1.2.3.4.6.7.8-HpCDD [c]	17.4 (12.4 - 30.1)	17.8 (13.2 - 25.4)	23.4 (16.8 - 31.5)	0.30
OCDD [c]	93.1 (59.8 - 141.6)	89.6 (66.3 - 141.0)	105.9 (79.5 - 141.5)	0.62
Sum PCDDs [c]	157.1 (93.4 - 241.1)	136.9 (109.8 - 221.5)	167.1 (149.0 - 238.1)	0.33
2.3.7.8-TCDF [c]	1.2 (0.6 - 2.1)	1.3 (1.0 - 1.6)	1.8 (1.1 - 2.5)	0.19
1.2.3.7.8-PeCDF [c]	0.7 (0.4 - 1.3)	0.9 (0.7 - 1.0)	1.1 (0.6 - 1.4)	0.26
2.3.4.7.8-PeCDF [c]	15.1 (8.8 - 26.0)	17.8 (10.7 - 20.8)	21.6 (13.2 - 37.5)	0.13
1.2.3.4.7.8-HxCDF [c]	4.6 (3.4 - 7.1)	5.6 (3.9 - 7.2)	5.6 (4.3 - 9.7)	0.25
1.2.3.6.7.8-HxCDF [c]	4.6 (3.2 - 7.0)	5.9 (3.7 - 6.8)	6.4 (4.3 - 9.9)	0.23
1.2.3.7.8.9-HxCDF [c]	0.3 (0.1 - 0.4)	0.3 (0.2 - 0.4)	0.3 (0.2 - 0.5)	0.28
2.3.4.6.7.8- HxCDF [c]	2.5 (1.5 - 3.8)	2.5 (1.7 - 3.3)	2.6 (1.7 - 4.2)	0.50
1.2.3.4.6.7.8-HpCDF [c]	4.0 (2.3 - 5.1)	3.9 (3.4 - 6.1)	4.0 (2.8 - 6.2)	0.70
1.2.3.4.7.8.9-HpCDF [c]	0.3 (0.3 - 0.6)	0.4 (0.3 - 0.6)	0.4 (0.3 - 0.5)	0.53
OCDF [c]	0.8 (0.6 - 1.2)	0.9 (0.5 - 1.5)	1.0 (0.9 - 1.6)	0.17
Sum PCDFs [c]	35.0 (20.9 - 54.9)	41.0 (32.1 - 49.2)	48.0 (31.1 - 70.4)	0.19
OMS-TEQ PCDD/F [d]	15.6 (11.4 - 27.6)	19.0 (13.9 - 22.7)	25.7 (14.9 - 36.5)	0.12
PCB 77 [c]	9.4 (7.4 - 14.0)	11.7 (7.7 - 19.5)	14.5 (9.4 - 23.4)	0.13
PCB 81 [c]	5.3 (3.0 - 9.4)	4.8 (3.0 - 7.1)	7.4 (4.7 - 12.3)	0.31
PCB 126 [c]	77.5 (40.5 - 174.4)	83.2 (50.7 - 146.6)	123.2 (100.8 - 314.9)	0.035
PCB 169 [c]	51.6 (28.4 - 99.8)	49.3 (34.6 - 58.8)	88.2 (43.1 - 136.6)	0.06
Sum Copl. PCBs [c]	142.2 (78.3 - 317.2)	148.3 (106.1 - 243.0)	222.7 (170.6 - 502.9)	0.035
PCB 105 [c]	5001.6 (2434.2 - 10041.8)	5385.0 (3028.2 - 9069.6)	8058.6 (6816.1 - 14580.4)	0.046
PCB 114 [c]	1177.5 (481.4 - 2158.6)	950.8 (674.7 - 1549.6)	1892.2 (1116.4 - 3229.9)	0.039
PCB 118 [c]	22777.3 (10346.1 - 44823.8)	20028.5 (12433.6 - 35803.7)	38553.2 (27118.5 - 71359.0)	0.039
PCB 123 [c]	240.5 (119.0 - 524.5)	241.4 (119.1 - 468.7)	444.9 (334.8 - 842.8)	0.033
PCB 156 [c]	9208.5 (4182.1 - 18217.1)	8799.5 (5088.2 - 10683.7)	15913.1 (7700.4 - 20898.5)	0.059
PCB 157 [c]	1743.5 (687.5 - 3653.1)	1353.4 (835.1 - 1844.1)	2923.9 (1329.0 - 3435.5)	0.026
PCB 167 [c]	2422.3 (1117.8 - 6153.4)	2203.1 (1611.9 - 3375.9)	4658.6 (2565.7 - 6980.3)	0.025
PCB 189 [c]	873.4 (408.8 - 1790.9)	786.8 (539.5 - 1000.3)	1855.5 (665.4 - 2200.5)	0.079
Sum Non Cop. PCBs [c]	42336.3 (19966.5 - 90500.2)	36028.2 (25865.5 - 65409.0)	77086.9 (48508.3 - 144809.2)	0.029
OMS-TEQ dl-PCB [d]	10.0 (5.2 - 24.2)	10.3 (7.2 - 18.6)	15.7 (13.7 - 41.0)	0.038
PCB 28 [e]	3.8 (2.4 - 6.6)	3.7 (2.1 - 8.0)	5.0 (2.9 - 7.5)	0.69
PCB 52 [e]	0.8 (0.5 - 1.0)	0.8 (0.5 - 1.4)	0.9 (0.7 - 1.5)	0.44
PCB 101 [e]	1.3 (0.8 - 1.7)	1.2 (0.8 - 2.2)	1.9 (1.4 - 2.1)	0.10
PCB 138 [e]	50.5 (23.3 - 121.9)	59.6 (36.9 - 82.4)	97.9 (56.2 - 145.1)	0.050
PCB 153 [e]	107.2 (47.4 - 229.1)	111.6 (74.0 - 141.7)	197.8 (97.7 - 256.8)	0.10
PCB 180 [e]	60.4 (29.3 - 125.6)	61.6 (39.7 - 70.4)	124.5 (47.2 - 184.2)	0.062
Sum 6 ndl-PCB [e]	221.5 (100.0 - 484.3)	231.1 (162.5 - 313.3)	456.1 (203.2 - 578.0)	0.074

PBDE 28 [e]	0.11 (0.05 - 0.15)	0.09 (0.06 - 0.15)	0.11 (0.08 - 0.16)	0.56
PBDE 47 [e]	1.30 (0.64 - 1.79)	1.27 (0.77 - 2.23)	1.22 (0.88 - 2.54)	0.48
PBDE 99 [e]	0.24 (0.15 - 0.40)	0.27 (0.17 - 0.40)	0.24 (0.20 - 0.48)	0.60
PBDE 100 [e]	0.34 (0.14 - 0.52)	0.34 (0.17 - 0.59)	0.34 (0.25 - 0.76)	0.50
PBDE 153 [e]	1.76 (0.95 - 2.65)	1.38 (0.95 - 2.62)	2.46 (1.73 - 3.03)	0.12
PBDE 154 [e]	0.05 (0.03 - 0.09)	0.08 (0.05 - 0.19)	0.07 (0.04 - 0.09)	0.18
PBDE 183 [e]	0.14 (0.09 - 0.23)	0.16 (0.10 - 0.22)	0.18 (0.14 - 0.30)	0.20
PBDE 209 [e]	0.87 (0.23 - 1.77)	0.52 (0.11 - 2.41)	0.40 (0.05 - 1.21)	0.25
Sum 7 i PBDE [e]	4.34 (2.17 - 5.57)	4.63 (2.45 - 7.34)	4.94 (3.88 - 6.76)	0.39
PBB 153 [e]	0.22 (0.11 - 0.46)	0.18 (0.14 - 0.23)	0.25 (0.17 - 0.66)	0.037

Units: [a] Kcal/kg bw /day; [b] g/kg bw /day; [c] pg/kg bw /day; [d] TEF/kg bw /day; [e] ng/kg bw /day

Figure 5. Forest plot representing the associations (β coefficient, 95% confidence interval) between postnatal exposure of preterm infants to POPs (continuous variable, log-scale), through breastmilk, and growth during hospital stay, measured as the difference of weight Z-score between birth and discharge (continuous variable) (n= 68). **Model 1**, crude model. **Model 2**, adjusted for birth weight, hospital stay duration, maternal age, maternal body mass index and infant energy intake. Statistical significance is represented of coefficients is represented by * p<0.05 and ** p<0.01.

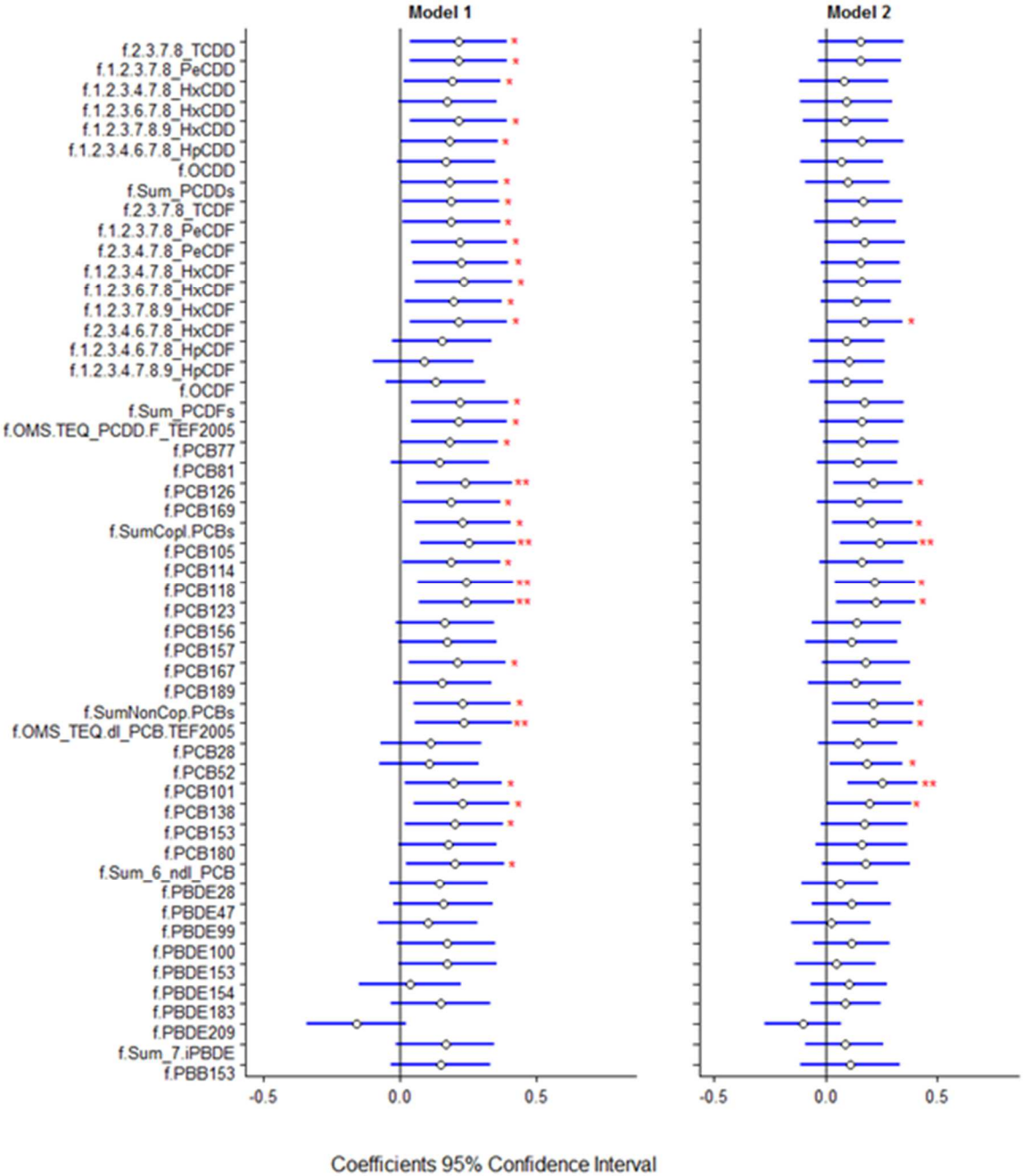
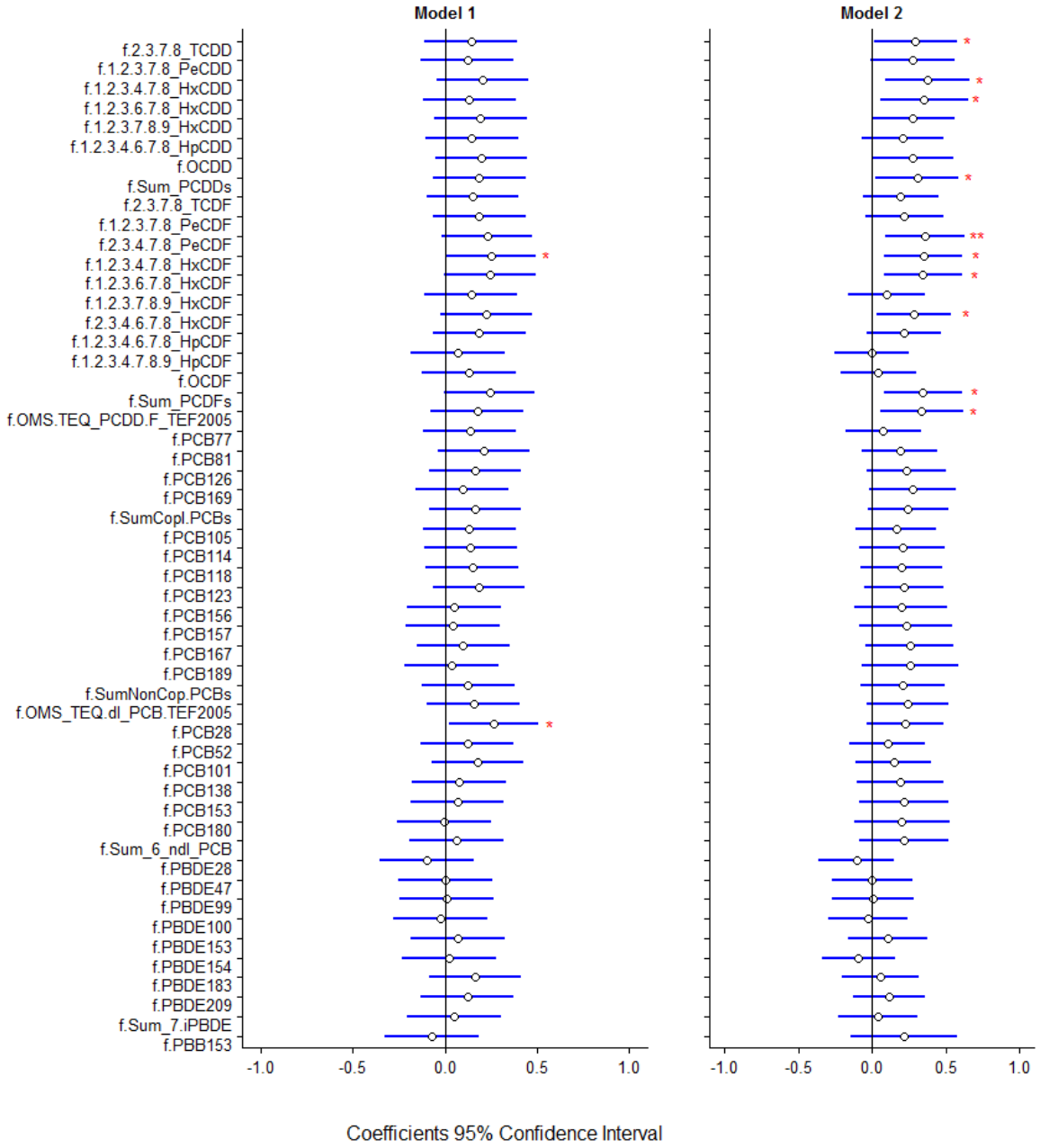


Figure 6. Forest representing the associations (β coefficient, 95% confidence interval) between exposure of preterm infants to POPs (continuous variable, log-scale) through breastmilk and weight z-score at 24 months (continuous variable) (n= 54). **Model 1**, crude model. **Model 2**, adjusted for birth weight, infant energy intake during delivery hospital stay, maternal age and maternal BMI. Statistical significance of coefficients is represented by * p<0.05 and ** p<0.01.



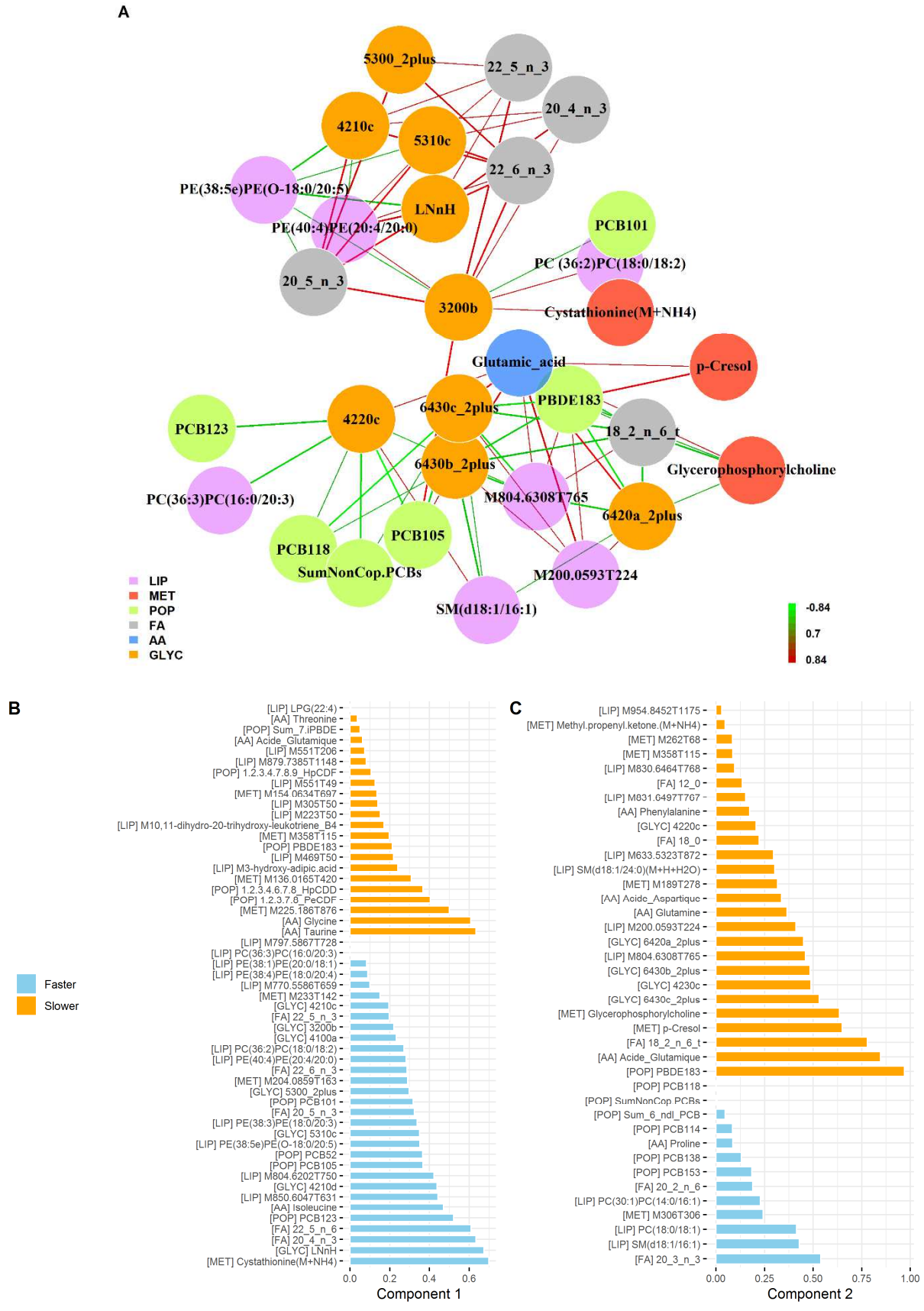
3.4. Integrated expotypes and nutritypes associated to preterm infant growth

After the iterative tuning process conducted within DIABLO framework, we finally selected a null design for the correlation matrix that emphasized the selection process of most discriminatory variables and maximized the separation of growth groups. The CV optimization process led us to

setup the model with 2 components. We established, for the blocks lipidome, metabolome, glycome, POPs, FAs and AAs, a list of 20, 7, 7, 9, 5 and 5 variables selected, respectively in the first component and 10, 7, 5, 7, 5 and 5 variables selected for the second component. The scatter plots from the multiblock model showed the most discriminant ability for the blocks lipidome, metabolome and POPs (Figure S2). Among the unknown molecules retained from the non-targeted approaches, we were able to finally annotate 14 lipids, including five phosphatidylethanolamines (plasmalogen-derivative PE PE(38:5e): PE(O-18:0/20:5), PE(38:3): PE(18:0/20:3), PE(40:4): PE(20:4/20:0), PE(38:4): PE(18:0/20:4), PE(38:1): PE(20:0/18:1)), three phosphatidylcholines (PC(36:2): PC(18:0/18:2), PC(42:1): PC(18:1/24:0) and PC(30:1): PC(14:0/16:1)) , two sphingomyelins (SM (d18:1/16:1) and SM(d18:1/24:0)), one diacylglycerol (22:5n-3/22:6n-3), one fatty acyl (3-hydroxy-adipic acid), one lysophosphatidylglycerol (LPG(22:4)), and one leukotriene (M10,11-dihydro-20-trihydroxy-leukotriene-B4). Additionally, we identified 3 molecules from the metabolome dataset (cystathionine, glycerophosphorylcholine and p-cresol). The associations between variables across blocks can be visualized in the correlation individual plots (Figure S3A and S3B) and in the correlation between molecular variables in the circle plots (Figure S3C and S3D) for the component 1 and 2, respectively. The circle plot from the component 1 highlighted mainly positive correlations between lipids with oligosaccharides and some POPs, conversely the component 2 showed a higher number of negative correlations between biomarkers, specially between glycome, lipidome datasets, some POPs and the glutamic acid. The relevance network plot displayed in the Figure 7A highlights the most relevant clusters of variables that discriminated the preterm infant growth groups, supported by the clustered image map (Figure S4). The loading vectors for each selected variable from each molecular block is illustrated in the Figure S5, grouped by blocks, and in the Figure 7B and 7C, grouped by the growth class where the molecules are more abundant, sorted by the loading weights. Globally, the chemical signature observed for the “Faster growth” group exhibited a characteristic profile of polyunsaturated fatty acids type ω-3 and ω-6, phosphatidylethanolamines, phosphatidylcholines, the branched chain amino acid isoleucine, the oligosaccharide Lacto-N-neohexaose (LNnH) and the metabolite cystathionine. The “Slower growth” group was more associated with a molecular signature richer in the AAs: glycine, taurine and glutamic acid, the

metabolite p-cresol, some saturated fatty acids, the hydroxyl adipic acid and a leukotriene. We also observed a different pattern among groups in terms of POPs exposure, being the PBDEs (specially the PBDE183 and some dioxins) more associated with the slower growth group, whereas the PCBs (i.e. 123, 105, 52) were more abundant in the faster growth group.

Figure 7. 7.A. Relevance network plot representing the group of molecular variables more associated with the growth of preterm infants (cut-off of absolute correlation set-up at ± 0.7) generated by the multiblock model. The colour code of circles identifies the different bocks: metabolome (red), purple (lipidome), glycome (orange), amino acids (blue), fatty acids (grey) and persistent organic pollutants (green). 7.B. and 7.C. Bar plots representing the loading vectors from both components generated by the multiblock model, grouped by the growth class ("Faster" in blue; "Slower" in orange) where the molecules were more abundant. The block identifiers can be found in brackets: amino acids [AA], fatty acids [FA], glycome [GLYC], lipidome [LIP], metabolome [MET] and persistent organic pollutants [POP].


A
5300_2plus
22_5_n_3
20_4_n_3
4210c
5310c
22_6_n_3
PE(38:5e)PE(O-18:0/20:5)
LNnH
PE(40:4)PE(20:4/20:0)
PCB101
PC (36:2)PC(18:0/18:2)
20_5_n_3
3200b
Cystathionine(M+NH4)
Glutamic_acid
p-Cresol
PBDE183
6430c_2plus
4220c
PCB123
18_2_n_6_t
6430b_2plus
Glycerophosphorylcholine
PC(36:3)PC(16:0/20:3)
M804.6308T765
6420a_2plus
PCB118
PCB105
SumNonCop.PCBs
M200.0593T224
SM(d18:1/16:1)
LIP
MET
POP
FA
AA
GLYC
-0.84
0.7
0.84
B
[LIP] LPG(22:4)
[AA] Threonine
[POP] Sum_7.iPBDE
[AA] Acide_Glutamique
[LIP] M551T206
[LIP] M879.7385T1148
[POP] 1.2.3.4.7.8.9_HpCDF
[LIP] M551T49
[MET] M154.0634T697
[LIP] M305T50
[LIP] M223T50
[LIP] M10,11-dihydro-20-trihydroxy-leukotriene_B4
[MET] M358T115
[POP] PBDE183
[LIP] M469T50
[LIP] M3-hydroxy-adipic.acid
[MET] M136.0165T420
[POP] 1.2.3.4.6.7.8_HpCDD
[POP] 1.2.3.7.8_PeCDF
[MET] M225.186T876
[AA] Glycine
[AA] Taurine
[LIP] M797.5867T728
[LIP] PC(36:3)PC(16:0/20:3)
[LIP] PE(38:1)PE(20:0/18:1)
[LIP] PE(38:4)PE(18:0/20:4)
[LIP] M770.5586T659
[MET] M233T142
[GLYC] 4210c
[FA] 22_5_n_3
[GLYC] 3200b
[GLYC] 4100a
[LIP] PC(36:2)PC(18:0/18:2)
[LIP] PE(40:4)PE(20:4/20:0)
[FA] 22_6_n_3
[MET] M204.0859T163
[GLYC] 5300_2plus
[POP] PCB101
[FA] 20_5_n_3
[LIP] PE(38:3)PE(18:0/20:3)
[GLYC] 5310c
[LIP] PE(38:5e)PE(O-18:0/20:5)
[POP] PCB52
[POP] PCB105
[LIP] M804.6202T750
[GLYC] 4210d
[LIP] M850.6047T631
[AA] Isoleucine
[POP] PCB123
[FA] 22_5_n_6
[FA] 20_4_n_3
[GLYC] LNnH
[MET] Cystathionine(M+NH4)
0.0
0.2
0.4
0.6
Component 1
Faster
Slower
C
[LIP] M954.8452T1175
[MET] Methyl.propenyl.ketone.(M+NH4)
[MET] M262T68
[MET] M358T115
[LIP] M830.6464T768
[FA] 12_0
[LIP] M831.6497T767
[AA] Phenylalanine
[GLYC] 4220c
[FA] 18_0
[LIP] M633.5323T872
[LIP] SM(d18:1/24:0)(M+H+H2O)
[MET] M189T278
[AA] Acide_Aspartique
[AA] Glutamine
[LIP] M200.0593T224
[GLYC] 6420a_2plus
[LIP] M804.6308T765
[GLYC] 6430b_2plus
[GLYC] 4230c
[GLYC] 6430c_2plus
[MET] Glycerophosphorylcholine
[MET] p-Cresol
[FA] 18_2_n_6_t
[AA] Acide_Glutamique
[POP] PBDE183
[POP] PCB118
[POP] SumNonCop.PCBs
[POP] Sum_6_ndl_PCB
[POP] PCB114
[AA] Proline
[POP] PCB138
[POP] PCB153
[FA] 20_2_n_6
[LIP] PC(30:1)PC(14:0/16:1)
[MET] M306T306
[LIP] PC(18:0/18:1)
[LIP] SM(d18:1/16:1)
[FA] 20_3_n_3
0.00
0.25
0.50
0.75
1.00
Component 2

4. Discussion

The literature has previously shown that *in utero* exposure to some organochlorine pesticides including dichlorodiphenyldichloroethylene (*p,p'*-DDE) or chlordecone may be associated with preterm birth (Longnecker et al., 2001). The associations with other POPs, such as PCDD/Fs or PCBs were reported as insignificant (Casas et al., 2015; Ferguson and Chin, 2017; Longnecker et al., 2005). Nonetheless, one study has shown significant positive associations between exposure to PBDEs and the risk preterm birth (Peltier et al., 2015). To the best of our knowledge, no studies have explored the impact of postnatal exposure to POPs on early growth of preterm infants, highlighting the value of our small cohort. The nutritional interventions with fortified breastmilk are recommended to favour the high energy and protein intake required to achieve the normal postnatal growth and neurodevelopment of preterm infant (Gibertoni et al., 2015). Considering the internal exposure bio-amplification among low and very low birth weight infants, we hypothesized that the xenobiotic component of breastmilk, specifically lipophilic pollutants, could contribute to impair the normal preterm growth interacting with the nutritional fraction.

To this end, we have applied for the first time a novel comprehensive approach based on MS multi-platform coupled to an advanced data workflow to better understand the molecular complexity of breastmilk and its simultaneous impact on preterm infant growth. The analytical approach includes a combination of ultra-trace targeted methods to quantify accurately low abundance contaminants (i.e. dioxins) combined with global metabolomics and lipidomics profiling methods to extend the framework to capture potential nutrient-contaminant associations. The potential of metabolomics in epidemiological exposome-health studies is vast. Some examples include the identification of endogenous metabolites to gain insights into the underlying pathological processes or to better understand the role of exposure biomarkers and molecular mediators, informing about specific exposure-to-disease

pathways (Assi et al., 2015; Chadeau-Hyam et al., 2011; Niedzwiecki et al., 2019). Additionally, the development of non-targeted metabolomics to identify exogenous stressors in human samples at a massive scale is growing rapidly (Andra et al., 2017; Dennis et al., 2017; Jones, 2016; Warth et al., 2017), shaping a novel path for environmental epidemiology. Actually, the high-throughput generation of exposure biomarkers has been already demonstrated as a promising tool to support data-driven approaches, however, a large sample size is imperative to statistically identify the subtle associations commonly found between environmental exposures and diseases with enough statistical power after correcting for the false discovery rate (Chung et al., 2019; Manrai et al., 2017). Currently, there is not consensus about the best approach (i.e. data-driven vs hypothesis-driven) to identify complex associations between chemical mixtures and health outcomes. On the one hand, data-driven or comprehensive approaches benefit of a high discovery potential by screening a large panel of molecules at the cost of correction for the false discovery rates. On the other hand, hypothesis-driven approaches takes advantage of available knowledge to depict more biologically consistent models. For that reason, we believe that combining ultra-trace methods with semi-target metabolomics may provide a technical compromise between accuracy and selectivity to extend hypothesis-driven approaches towards a more comprehensive paradigm.

The results from the linear models, built on 58 mother-infant dyads, between the postnatal exposure to POPs and growth outcomes did not support an impairment of preterm infant growth during the hospital stay (Delta weight z-score between birth and discharge) or at 24 months (weight z-score). Conversely, we observed statistically significant and positive associations between some PCBs and delta weight z-score strongly influenced by the adjustment by energy intake. Similarly, positive associations were observed between some dioxins and the weight z-score at 24 months. The associations were no longer significant after

120 the FDR correction. The positive associations could be strongly confounded by the energy
121 and fat intake of infants considering the limitations to accurately estimate the energy and milk
122 intake among breastfeeding new-borns, especially during the time after the hospital stay. The
123 lack of concordance between the growth models at discharge and 24 months could be
124 explained by the limited information on infants' dietary habits and lifestyle potentially
125 confounding the associations. Another hypothesis could be the obesogenic potential of some
126 PCBs, for instance PCB-77 has been associated to increased adipogenesis and gain of body
127 fat (Arsenescu et al., 2008). However, this was not supported by the results at 24 months,
128 when the group of dioxins (inhibitors of adipogenic differentiation) positively associated with
129 weight z-score (Hsu et al., 2010). Moreover, it is necessary to keep in mind that these
130 expotypes and nutritypes signatures may be very different in case of full-term newborn
131 infants. Indeed, preterm infants, particularly male infants, are reported to experience
132 insufficient fat-free mass accretion but higher fat mass percentage during hospitalization in
133 neonatal intensive care unit (Simon et al, 2013) which likely interfere with lipophilic
134 pollutants exposure. In any case, further research should extend the present study, especially
135 with a larger population (preterm *versus* term), through a longer follow-up and gaining
136 insight about the metabolic status and lean-fat mass ratio of children, taking into account their
137 sex-related differences.
138 For the first time, we have shown with an integrative multi-platform approach the complex
139 exposome-metabolome chemical space of human milk and its potential interactive impact on
140 preterm outcomes. For instance, the milks from the 'slower' growth group presented a
141 characteristic profile of saturated fatty acids whereas the 'faster' group were mainly
142 polyunsaturated. The previous analysis of preterm milk reduced to the lipidomic block alone
143 (Alexandre-Gouabau et al, 2018) highlighted the major effect of several ω-3 FAs, such as
144 DHA and its precursors EPA and DPA on postnatal infant growth during the hospital stay.

We also noticed high correlation between some PCBs and ω-3 FAs found in the analysed breastmilk, reflecting the common dietary pathways such as fish and seafood products. Interestingly, some studies have found that ω-3 FAs could counterbalance the harmful effects of POPs (Marushka et al., 2017). This interaction could also explain the inconsistent results on the beneficial effects of dietary intake of ω-3 FAs for cognitive function and growth of preterm infants that have been reported (Smith and Rouse, 2017). Interestingly, using the multiblock model we identified PBDE183 and some dioxins (12378-PeCDF and 1234678-HpCDD), being more associated with the "slower" growth group. These chemicals were not highlighted by the linear growth models; thus, it may suggest that molecular covariation with the rest of molecules could mask an elusive underlying association. The impact of negative or positive confounding of nutrients in environmental epidemiology has been scarcely investigated, however the few studies that accounted for it has reported interesting findings. For instance the negative confounding was elucidated with methyl mercury (MetHg) and polyunsaturated fatty acids on the associations with neurotoxicity of children from Faroe Islands, well known by their high fish intake (Choi et al., 2014). The co-exposure of MetHg, omega-3 fatty acids and selenium has been reported also in cord blood, resulting in negative confounding and effect modification on the associations with foetal growth outcomes (Wells et al., 2016).

The higher nutritional content related with "faster" growth group is consistent with anabolic function associated to the specific identified nutrients and related metabolites, such as, for example, the higher abundance of isoleucine, a branched-chain amino acid, in the "faster" milks, an amino acid previously associated with a higher early weight gain (Alexandre-Gouabau et al., 2019; Kirchberg et al., 2015). Conversely, as in the previous preterm milk metabolomics analysis without considering pollutants, we observed higher abundance of taurine and glycine, under their free form, in "slower" breast milk along with a lower total fat

content. Taurine contributes to intestinal fat absorption, bile acid secretion and hepatic functions, but no associations have been reported with growth outcomes among preterm or low birth weight infants in randomized trials (Verner et al., 2007). The present results could be explained by a depleted bile acid conjugation by taurine and glycine in "slower" breast milk, that could subsequently limit the solubilisation of lipids and sterols in mixed micelles, resulting, in turn, in the limited uptake of lipids into enterocytes (Alexandre-Gouabau et al., 2019) in addition to preterm gastrointestinal immaturity. The extend of the impact of milk composition on infant's absorption of organochlorines remains largely unexplored, being most of pharmacokinetic models developed under very simplistic assumptions (Lehmann et al., 2014).

The nutritional metabolites identified in the present study were slightly different from our previous analysis without considering POPs, probably explained by multi-block approach used and the potential covariation due to the block of POPs (Alexandre-Gouabau et al., 2019). Actually, the development and application of multiblock models is in full emergence, especially in the field of molecular biology or clinical settings where the cross-omic and multi-way integration is more common (Sun and Hu, 2016; Wu et al., 2019). In the present study, we have applied a powerful statistical approach that allows the variable selection accounting for the correlation between variables but also between data-blocks, presumably outperforming the concatenation approaches (Singh et al., 2019). This proposed approach may extend the growing toolkit to analyse the complex exposome-health data structures, favouring the holistic analysis of multiple exposures and potential mediators (Lazarevic et al., 2019). Additionally, it could be easily coupled to generic physiologically based pharmacokinetic (PBPK) models for POPs to refine the exposure estimates through different exposure windows or to generate cumulative exposure signatures through the perinatal period (Verner et al., 2013).

5. Conclusions

To sum up, the results from the present study support the methodological potential of cross-platforms integration to elucidate complex exposome-health associations. Despite we applied the workflow to identify nutritional metabolites in human preterm breastmilk, this could be applied to other matrices such as serum or adipose tissue to identify endogenous metabolites favouring the functional phenotyping. Furthermore, these preliminary results did not elucidate an impairment of preterm infant growth associated to the postnatal exposure to POPs, supporting the use of human milk in the nutritional strategies for preterm newborns. However, further research will be required 1) to evaluate the associations with higher exposure concentrations of POPs and more complex mixtures with emerging contaminants, 2) to gain insight into the suggested nutrient-pollutants interactions and 3) to evaluate the potential obesogenic effect of POPs on preterm infants that could exacerbate their already acknowledged risk for metabolic diseases (Chehade et al., 2018).

6. Author contributions

C.-Y.B., J.-C.R.: conceived the LACTACOL study and design. C.-Y.B., B.L.B and J.P.A conceived the LACTAPOP study on a subset of samples. A.-L.R., Y.G., F.F., S.C.: conducted metabolomics analysis and/or metabolites identification. C.-Y.B., J.-C.R, H.B. managed the clinical data, including collection and data cleaning. M.-C.A.-G.: conceived and supervised the present metabolomics study conducted on a subset of infants of LACTACOL cohort; G.C.S and T.M. conceived the appropriate statistical tools and/or performed the statistical analysis; G.C.S conducted the statistical analysis and provided a first draft; M.-C.A.-G ; J. P.A, B.L.B, T.M., F.F., C.-Y.B. and D.D.: critically revised the manuscript for important intellectual content; and all authors: acquired, analyzed, or interpreted the data, and read and approved the final manuscript.

7. Acknowledgements

The LACTACOL projet was funded by Région Pays-de-la-Loire (2011–2012) and by the ‘Fonds Européen de Développement Economique et Régional» (FEDER, Grant 38395, Project 6226, 2013–2015).

8. References

Agostoni C, Buonocore G, Carnielli VP, De Curtis M, Darmaun D, Decsi T, et al. Enteral nutrient supply for preterm infants: commentary from the European Society of Paediatric Gastroenterology, Hepatology and Nutrition Committee on Nutrition. J Pediatr Gastroenterol Nutr 2010; 50: 85-91.

Alexandre-Gouabau M-C, Moyon T, David-Sochard A, Fenaille F, Cholet S, Royer A-L, et al. Comprehensive Preterm Breast Milk Metabotype Associated with Optimal Infant Early Growth Pattern. Nutrients 2019; 11: 528.

Alexandre-Gouabau MC, Moyon T, Cariou V, Antignac JP, Qannari EM, Croyal M, et al. Breast Milk Lipidome Is Associated with Early Growth Trajectory in Preterm Infants. Nutrients 2018; 10.

Andra SS, Austin C, Patel D, Dolios G, Awawda M, Arora M. Trends in the application of high-resolution mass spectrometry for human biomonitoring: An analytical primer to studying the environmental chemical space of the human exposome. Environ Int 2017; 100: 32-61.

Antweiler RC. Evaluation of Statistical Treatments of Left-Censored Environmental Data Using Coincident Uncensored Data Sets. II. Group Comparisons. Environ Sci Technol 2015; 49: 13439-46.

Arsenescu V, Arsenescu RI, King V, Swanson H, Cassis LA. Polychlorinated biphenyl-77 induces adipocyte differentiation and proinflammatory adipokines and promotes obesity and atherosclerosis. Environ Health Perspect 2008; 116: 761-8.

Assi N, Fages A, Vineis P, Chadeau-Hyam M, Stepien M, Duarte-Salles T, et al. A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. Mutagenesis 2015; 30: 743-53.

Boquien CY. Human Milk: An Ideal Food for Nutrition of Preterm Newborn. Front Pediatr 2018; 6: 295.

247 Boue G, Cummins E, Guillou S, Antignac JP, Le Bizec B, Membre JM. Public health risks and
248 benefits associated with breast milk and infant formula consumption. Crit Rev Food Sci Nutr
249 2018; 58: 126-145.
250 Casas M, Nieuwenhuijsen M, Martinez D, Ballester F, Basagana X, Basterrechea M, et al. Prenatal
251 exposure to PCB-153, p,p'-DDE and birth outcomes in 9000 mother-child pairs: exposure-
252 response relationship and effect modifiers. Environ Int 2015; 74: 23-31.
253 Chadeau-Hyam M, Athersuch TJ, Keun HC, De Iorio M, Ebbels TM, Jenab M, et al. Meeting-in-the-
254 middle using metabolic profiling - a strategy for the identification of intermediate biomarkers
255 in cohort studies. Biomarkers 2011; 16: 83-8.
256 Chehade H, Simeoni U, Guignard J-P, Boubred F. Preterm Birth: Long Term Cardiovascular and
257 Renal Consequences. Current pediatric reviews 2018; 14: 219-226.
258 Choi AL, Mogensen UB, Bjerve KS, Debes F, Weihe P, Grandjean P, et al. Negative confounding by
259 essential fatty acids in methylmercury neurotoxicity associations. Neurotoxicol Teratol 2014;
260 42: 85-92.
261 Chun H, Keleş S. Sparse partial least squares regression for simultaneous dimension reduction and
262 variable selection. Journal of the Royal Statistical Society. Series B, Statistical methodology
263 2010; 72: 3-25.
264 Chung MK, Buck Louis GM, Kannan K, Patel CJ. Exposome-wide association study of semen
265 quality: Systematic discovery of endocrine disrupting chemical biomarkers in fertility require
266 large sample sizes. Environ Int 2019; 125: 505-514.
267 Commission Regulation N° 252/2012/EC Laying down methods of sampling and analysis for the
268 official control of levels of dioxins, dioxin- like PCBs and nondioxin- like PCBs in certain
269 foodstuffs and repealing Regulation (EC) No 1883/2006.
270 Courant F, Royer AL, Chéreau S, Morvan ML, Monteau F, Antignac JP, et al. Implementation of a
271 semi-automated strategy for the annotation of metabolomic fingerprints generated by liquid
272 chromatography-high resolution mass spectrometry from biological samples Analyst 2012,
273 137, 4958-4967.

Dennis KK, Marder E, Balshaw DM, Cui Y, Lynes MA, Patti GJ, et al. Biomonitoring in the Era of the Exposome. Environ Health Perspect 2017; 125: 502-510.

Ferchaud-Roucher V, Croyal M, Krempf M, Ouguerram K. Plasma lipidome characterization using UHPLC-HRMS and ion mobility of hypertriglyceridemic patients on nicotinic acid. Atherosclerosis 2015; 241: e123-e124.

Ferguson KK, Chin HB. Environmental chemicals and preterm birth: Biological mechanisms and the state of the science. Curr Epidemiol Rep 2017; 4: 56-71.

Ghaziri A, Qannari EM, Moyon T, Alexandre-Gouabau M-C. AoV-PLS: A new method for the analysis of multivariate data depending on several factors. 2015; 8: 214-235.

Gibertoni D, Corvaglia L, Vandini S, Rucci P, Savini S, Alessandroni R, et al. Positive effect of human milk feeding during NICU hospitalization on 24 month neurodevelopment of very low birth weight infants: an Italian cohort study. PLoS One 2015; 10: e0116552.

Hiltunen H, Loyttyniemi E, Isolauri E, Rautava S. Early Nutrition and Growth until the Corrected Age of 2 Years in Extremely Preterm Infants. Neonatology 2018; 113: 100-107.

Hsu HF, Tsou TC, Chao HR, Kuo YT, Tsai FY, Yeh SC. Effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on adipogenic differentiation and insulin-induced glucose uptake in 3T3-L1 cells. J Hazard Mater 2010; 182: 649-55.

Jones DP. Sequencing the exposome: A call to action. Toxicol Rep 2016; 3: 29-45.

Kirchberg FF, Harder U, Weber M, Grote V, Demmelmair H, Peissner W, et al. Dietary protein intake affects amino acid and acylcarnitine metabolism in infants aged 6 months. J Clin Endocrinol Metab 2015; 100: 149-58.

Lazarevic N, Barnett AG, Sly PD, Knibbs LD. Statistical Methodology in Studies of Prenatal Exposure to Mixtures of Endocrine-Disrupting Chemicals: A Review of Existing Approaches and New Alternatives. Environ Health Perspect 2019; 127: 26001.

Lehmann GM, Verner M-A, Luukinen B, Henning C, Assimon SA, LaKind JS, et al. Improving the risk assessment of lipophilic persistent environmental chemicals in breast milk. Critical reviews in toxicology 2014; 44: 600-617.

301 Longnecker MP, Klebanoff MA, Brock JW, Guo X. Maternal levels of polychlorinated biphenyls in
302 relation to preterm and small-for-gestational-age birth. Epidemiology 2005; 16: 641-7.
303 Longnecker MP, Klebanoff MA, Zhou H, Brock JW. Association between maternal serum
304 concentration of the DDT metabolite DDE and preterm and small-for-gestational-age babies
305 at birth. Lancet 2001; 358: 110-4.
306 Manrai AK, Cui Y, Bushel PR, Hall M, Karakitsios S, Mattingly CJ, et al. Informatics and Data
307 Analytics to Support Exposome-Based Discovery for Public Health. Annu Rev Public Health
308 2017; 38: 279-294.
309 Marushka L, Batal M, David W, Schwartz H, Ing A, Fediuk K, et al. Association between fish
310 consumption, dietary omega-3 fatty acids and persistent organic pollutants intake, and type 2
311 diabetes in 18 First Nations in Ontario, Canada. Environ Res 2017; 156: 725-737.
312 Nauta MJ, Andersen R, Pilegaard K, Pires SM, Ravn-Haren G, Tetens I, et al. Meeting the challenges
313 in the development of risk-benefit assessment of foods. Trends in Food Science &
314 Technology 2018; 76: 90-100.
315 Niedzwiecki MM, Walker DI, Vermeulen R, Chadeau-Hyam M, Jones DP, Miller GW. The
316 Exposome: Molecules to Populations. Annu Rev Pharmacol Toxicol 2019; 59: 107-127.
317 Oursel S, Cholet S, Junot C, Fenaille F. Comparative analysis of native and permethylated human
318 milk oligosaccharides by liquid chromatography coupled to high resolution mass
319 spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 2017; 1071: 49-57.
320 Peltier MR, Koo HC, Getahun D, Menon R. Does exposure to flame retardants increase the risk for
321 preterm birth? J Reprod Immunol 2015; 107: 20-5.
322 Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: An R package for 'omics feature selection and
323 multiple data integration. PLoS Comput Biol 2017; 13: e1005752.
324 Simon L, Borrego P, Darmaun D, Legrand A, Rozé J, Chauty-Frondas A. Effect of sex and
325 gestational age on neonatal body composition. Brit J Nutr, 2013; 109:1105-1108.
326 Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, et al. DIABLO: an integrative
327 approach for identifying key molecular drivers from multi-omic assays. Bioinformatics 2019.

328 Smith SL, Rouse CA. Docosahexaenoic acid and the preterm infant. Matern Health Neonatol
329 Perinatol 2017; 3: 22.
330 Sun YV, Hu YJ. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of
331 Complex Human Diseases. Adv Genet 2016; 93: 147-90.
332 Tenenhaus A, Tenenhaus M. Regularized generalized canonical correlation analysis for multiblock or
333 multigroup data analysis. European Journal of Operational Research 2014; 238: 391-403.
334 Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal
335 Statistical Society: Series B (Statistical Methodology) 2011; 73: 273-282.
336 van den Berg M, Kypke K, Kotz A, Tritscher A, Lee SY, Magulova K, et al. WHO/UNEP global
337 surveys of PCDDs, PCDFs, PCBs and DDTs in human milk and benefit-risk evaluation of
338 breastfeeding. Arch Toxicol 2017; 91: 83-96.
339 Verner A, Craig S, McGuire W. Effect of taurine supplementation on growth and development in
340 preterm or low birth weight infants. Cochrane Database Syst Rev 2007: Cd006072.
341 Verner MA, Sonneborn D, Lancz K, Muckle G, Ayotte P, Dewailly E, et al. Toxicokinetic modeling
342 of persistent organic pollutant levels in blood from birth to 45 months of age in longitudinal
343 birth cohort studies. Environ Health Perspect 2013; 121: 131-7.
344 Warth B, Spangler S, Fang M, Johnson CH, Forsberg EM, Granados A, et al. Exposome-Scale
345 Investigations Guided by Global Metabolomics, Pathway Analysis, and Cognitive
346 Computing. Anal Chem 2017; 89: 11505-11513.
347 Wells EM, Herbstman JB, Lin YH, Jarrett J, Verdon CP, Ward C, et al. Cord Blood Methylmercury
348 and Fetal Growth Outcomes in Baltimore Newborns: Potential Confounding and Effect
349 Modification by Omega-3 Fatty Acids, Selenium, and Sex. Environ Health Perspect 2016;
350 124: 373-9.
351 WHO . Assessment of the health risk of dioxins: re-evaluation of the tolerable daily intake
352 (TDI).2000. Geneva, Switzerland, 25–29 May 1998. Food Addit Contam 17(4):223–369
353 WHO . Evaluation of certain food additives and contaminants. Geneva: World Health Organization;
354 2002.

355 Wu C, Zhou F, Ren J, Li X, Jiang Y, Ma S. A Selective Review of Multi-Level Omics Data
356 Integration Using Variable Selection. High Throughput 2019; 8.
357