

CLINICAL RESEARCH ARTICLE **OPEN**


# Metabolomic and exposomic biomarkers of risk of future neurodevelopmental delay in human milk

Kefeng Li<sup>1,2</sup>, Kerri Bertrand<sup>3</sup>, Jane C. Naviaux<sup>1,4</sup>, Jonathan M. Monk<sup>5</sup>, Alan Wells<sup>3</sup>, Lin Wang<sup>1,2</sup>, Sai Sachin Lingampelly<sup>1,2</sup>, Robert K. Naviaux<sup>1,2,3,6</sup>✉ and Christina Chambers<sup>3,7</sup>✉

© The Author(s) 2022

**BACKGROUND:** The chemical composition of human milk has long-lasting effects on brain development. We examined the prognostic value of the human milk metabolome and exposome in children with the risk of neurodevelopmental delay (NDD).

**METHODS:** This retrospective cohort study included 82 mother–infant pairs (40 male and 42 female infants). A total of 59 milk samples were from mothers with typically developing children and 23 samples were from mothers of children at risk. Milk samples were collected before 9 months of age ( $4.6 \pm 2.5$  months, mean  $\pm$  SD). Neurocognitive development was assessed by maternal report at  $14.2 \pm 3.1$  months using the Ages and Stages Questionnaires-2.

**RESULTS:** Metabolome and exposome profiling identified 453 metabolites and 61 environmental chemicals in milk. Machine learning tools identified changes in deoxysphingolipids, phospholipids, glycosphingolipids, plasmalogens, and acylcarnitines in the milk of mothers with children at risk for future delay. A predictive classifier had a diagnostic accuracy of 0.81 (95% CI: 0.66–0.96) for females and 0.79 (95% CI: 0.62–0.94) for males.

**CONCLUSIONS:** Once validated in larger studies, the chemical analysis of human milk might be added as an option in well-baby checks to help identify children at risk of NDD before the first symptoms appear.

*Pediatric Research* (2023) 93:1710–1720; <https://doi.org/10.1038/s41390-022-02283-6>

## IMPACT:

- Maternal milk for infants sampled before 9 months of age contained sex-specific differences in deoxysphingolipids, sphingomyelins, plasmalogens, phospholipids, and acylcarnitines that predicted the risk of neurodevelopmental delay at 14.2 months of age.
- Once validated, this early biosignature in human milk might be incorporated into well-baby checks and help to identify infants at risk so early interventions might be instituted before the first symptoms appear.

## INTRODUCTION

Neurodevelopmental delay (NDD) refers to the delays in skill development of infants and young children, including physical, linguistic, behavior, or learning areas.<sup>1</sup> According to a survey conducted in the United States for the years from 2009 to 2017, about 17% of children between the ages of 3–17 were affected by neurodevelopmental disorders, and recent trends have shown a significant increase in the prevalence of autism spectrum disorder since 2009.<sup>2</sup> Treatment of neurodevelopmental disorders can be difficult since some neurodevelopmental disorders cannot be confidently diagnosed until children are 2–3 years of age. Some symptoms, once developed, may continue throughout life, causing enormous social and economic burdens. Children diagnosed with neurodevelopmental disorders have better treatment outcomes if detected and treated at younger ages.<sup>3</sup>

Therefore, new methods for early prediction of an infant's later neurodevelopmental risk are urgently needed.

Early infancy is a critical period for neurodevelopment. As the ideal nutrition recognized by the World Health Organization for infants, human milk (HM) can have lasting effects on the structure and function of the developing brain. Compelling evidence supports the overall beneficial effects of HM intake on neurodevelopment.<sup>4–6</sup> However, limited information is available about the impact of milk composition on breastfed infants' neurocognitive outcomes.

Compositionally, human milk contains various amounts of protein, fat, carbohydrates, amino acids, and other metabolites, prebiotic factors like HM oligosaccharides, hormones, peptide growth factors, cytokines, microRNAs, minerals, and microbiota.<sup>7,8</sup> A recent study reported that protein content in HM is not

<sup>1</sup>The Mitochondrial and Metabolic Disease Center, University of California, San Diego School of Medicine, San Diego, CA, USA. <sup>2</sup>Department of Medicine, University of California, San Diego School of Medicine, San Diego, CA, USA. <sup>3</sup>Department of Pediatrics, University of California, San Diego School of Medicine, San Diego, CA, USA. <sup>4</sup>Department of Neurosciences, University of California, San Diego School of Medicine, San Diego, CA, USA. <sup>5</sup>Department of Bioengineering, University of California, San Diego School of Medicine, San Diego, CA, USA. <sup>6</sup>Department of Pathology, University of California, San Diego School of Medicine, San Diego, CA, USA. <sup>7</sup>Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego School of Medicine, San Diego, CA, USA. ✉email: rnaviaux@health.ucsd.edu; chchambers@health.ucsd.edu

Received: 24 January 2022 Revised: 25 May 2022 Accepted: 12 August 2022

Published online: 15 September 2022

correlated with preterm infant neurodevelopment at 2 years of age.<sup>9</sup> Metabolites in HM are the nutritional building blocks that support physical growth, immune system development, and brain maturation for infants.<sup>10,11</sup> The associations between metabolites in HM and the neurodevelopment and cognitive function in young offspring have only been sparsely examined and primarily focused on the total fat content and long-chain polyunsaturated fatty acid in HM.<sup>12,13</sup> For example, a recent systematic review has shown the association between fat content in HM for premature infants and neurodevelopmental outcomes.<sup>14</sup> In addition, higher linolenic acid and docosahexaenoic acid was found in milk collected at the third month of life were reported to be correlated with better psychomotor development of exclusively breastfed infants at the sixth month of life.<sup>15</sup> However, total fat concentration does not reflect different lipid classes and other endogenous metabolites in HM, which also appears to be important for infant growth.<sup>8</sup> Milk composition varies somewhat over the course of a day.<sup>16</sup> For example, the fat content varies from about 4.1–5.2% over a day,<sup>17</sup> and the time since the last feeding can influence the composition of milk available in the next feeding.<sup>18</sup> No studies have yet been published on the comprehensive analysis of the metabolomic differences in human milk between typically developing (TD) children and children with NDD.

A mother is exposed to various environmental chemicals through food, personal care products, household products, medical and recreational drugs, pollutants, or through her occupational environment. Some of this maternal body burden can be transferred to HM during lactation.<sup>19</sup> Infant early life exposure to environmental chemicals through milk is uniquely influential to their long-term outcomes due to heightened neuroplasticity during early childhood sensitive periods.<sup>20</sup> Experimentally, several common classes of environmental chemicals in HM have been assessed for their associations with neurodevelopmental outcomes of the offspring.<sup>21</sup> For example, in a mother–infant cohort from Vietnam, the neurodevelopmental scores for children at 4-month-old and 2-year-old were decreased with the increase in dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin) in HM collected at 1 month.<sup>22,23</sup> The level of diethylhexyl phthalate in HM collected at 30 days after delivery was associated with impaired mental development in children aged 1–2 years.<sup>24</sup> Scores for cognitive function were also lower in children who were exposed to higher amounts of the flame retardant brominated diphenyl ether-209 (BDE209) through HM.<sup>25</sup> However, most of the existing epidemiological studies for neurodevelopmental effects of HM have only focused on a small set of environmental chemicals and hence may exclude important chemicals from consideration. Since the general population can be simultaneously exposed to a wide range of environmental chemicals, a comprehensive characterization of the human milk exposome will allow an in-depth evaluation of the potential long-term neurodevelopmental effects of the HM exposome.

In this study, we conducted integrated metabolomic and exposomic profiling of HM. This broad-based approach combined with multiple machine learning (ML) algorithms aimed to (1) determine the potential metabolomic and exposomic differences in HM for infants with and without NDD risk, and (2) develop and evaluate multi-analyte models using HM metabolome and exposome data to predict infants with the future risk of NDD.

## MATERIALS AND METHODS

### Study participants and sample collection

The study population is based on the subjects from the University of California, San Diego (UC San Diego) Human Milk Biorepository, which contains over 5900 HM samples from over 2200 women who reside in the U.S. and Canada along with subsequent neurodevelopmental screening assessments for their infants/children. We first excluded maternal participants who reported diabetes, depression, other medical

conditions, or pregnancy complications. HM samples were collected using a standardized protocol previously described.<sup>26</sup> Inclusion criteria for this study were milk samples collected from mothers of infants less than 9 months of age and the valid completion of an Ages and Stages Questionnaire version 2 (ASQ2) during follow-up of the infant between 12 and 25 months of age. Exclusion criteria included two or more conflicting ASQ2 communication scores above and below the age-specific developmental threshold for a given child. HM samples were either collected at the participant's home and shipped on ice within 24 h of sample collection or collected on-site at UC San Diego and processed in the laboratory within 3 h of sample expression. Selected HM samples were aliquoted and stored at  $-80^{\circ}\text{C}$  before transfer to the Naviaux lab for analysis. All the study protocols were approved by the Institutional Review Board (IRB) at UC San Diego, and written consent was obtained upon the collection of the HM samples under Dr Chamber's UC San Diego IRB-approved protocol #130658. Metabolomics and exposomics were conducted on deidentified samples under Dr Naviaux's UC San Diego IRB-approved protocol #140072. Metabolomic and exposomic analysis of the samples was conducted from July 21, 2020, to September 22, 2020. Detailed participant and sample characteristics are summarized in Supplementary Table S1.

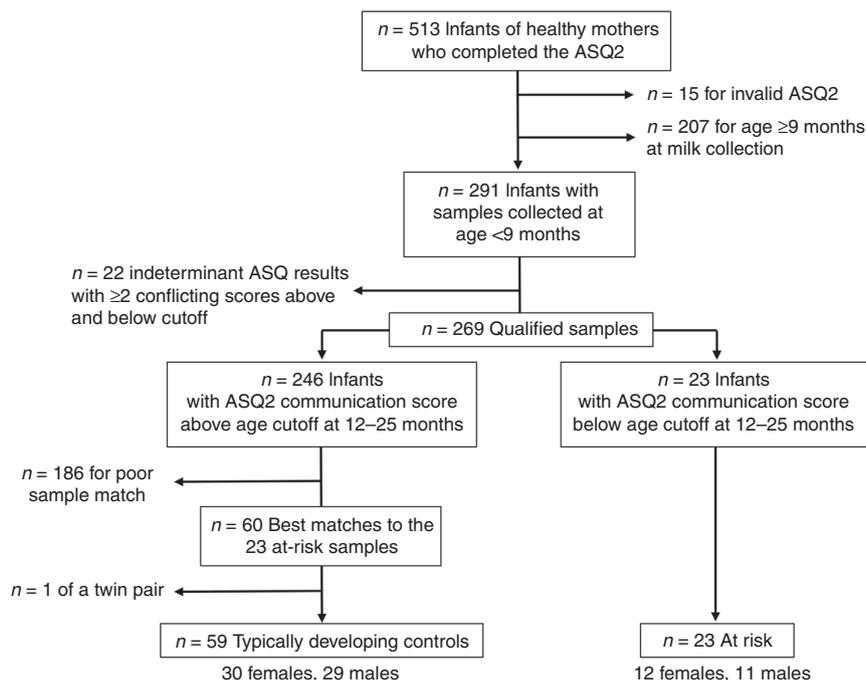
### Neurodevelopmental screening

Screening for neurocognitive abilities of the breastfed children was assessed by maternal reports using the ASQ2. ASQ2 is a parent/caregiver screening instrument that pinpoints developmental progress in children between the ages of one month and 5.5 years. The ASQ screening questionnaire has been found to have 84% overall agreement with validated tools such as the Bayley Scales of Infant Development.<sup>27,28</sup> The overall sensitivity of the screening instrument for correctly identifying children with delays and the specificity of the instrument for identifying TD children are 72 and 88%, respectively.<sup>27</sup> The questionnaire encompasses the domains of communication, gross motor, fine motor, problem solving, and personal social. Scores by domain are computed and compared to established cut-offs by child age and sex to identify areas of concern. For purposes of this study, children were classified as "screen positive for developmental delay" if they had an ASQ2 score that was critical or borderline in any single domain of the assessment. No genetic or biochemical data were available from the infants who tested at risk by ASQ2 score. Children were classified as "typically developing" (TD) if their scores in all domains were above the respective cutoff scores.

### Metabolomic and exposomic analysis

HM samples were thawed on ice and mixed with a set of internal standards. The endogenous metabolites and xenobiotic compounds were extracted with four volumes of 100% ethanol prechilled at  $-20^{\circ}\text{C}$ . Targeted, broad-spectrum, metabolomic analysis of endogenous metabolites was performed using a UFLC XR HPLC system (LC-20AD, Shimadzu) coupled with a Qtrap 5500 triple quadrupole mass spectrometer (SCIEX) (LC-MS/MS) as described earlier<sup>29,30</sup> with minor modifications to expand the list of targeted chemicals. To acquire comprehensive metabolomic profiles, we performed LC-MS/MS analysis in both hydrophilic interaction liquid chromatography (HILIC-MS/MS) mode and reverse phase (RP-MS/MS) mode. A total of 659 endogenous metabolites covering all major human metabolic pathways and diverse chemical classes were targeted. We used a Shodex polymer-based aminopropyl column (HILIC mode) for most endogenous metabolites and a Raptor biphenyl column (RP mode) for some of the lipophilic metabolites that cannot be measured well by the HILIC method, such as gangliosides, deoxyceramides, phosphatidylcholines, some sphingomyelins, and steroid hormones.

Exposome profiling of xenobiotic chemicals in HM was conducted using both RP-MS/MS and gas chromatography (GC)-MS/MS methods based on the chemical properties of the targets. A total of 937 anthropogenic compounds were targeted, including 145 analytes on the RP-MS/MS platform and 792 chemicals on the GC-MS/MS platform. These xenobiotics were selected to broadly monitor human exposure to commonly used pesticides, food additives, personal care products, pharmaceuticals, persistent organic pollutants, phthalates, and flame retardants in the US. MRM transitions and the retention time (RT) were optimized using the purified standards, and two MRMs transitions were used for each analyte. The full targeted list and the associated MS/MS parameters and RTs are listed in Supplementary Tables S2 and S3. See Supplementary Methods for additional details.



**Fig. 1 Participant and sample selection.** From the records of 513 milk samples stored as part of the UCSD Human Milk Biorepository, 269 infant-sample pairs met inclusion and exclusion criteria. From the milk samples collected before the infants were 9 months old, metabolomic and exposomic analysis was performed on 59 control and 23 at risk samples. Developmental screening was performed by ASQ screening at 1–2 years of age. ASQ2 Ages and Stages Questionnaire version 2.

### Comparison with the existing human milk metabolome data

To compare the HM endogenous metabolites detected in our study with known HM metabolome data, we downloaded the curated HM metabolites from the milk composition database and human metabolome database (HMDB). We also conducted an extensive literature review of known HM metabolites using the text-mining tool PolySearch.<sup>2,31</sup> All compounds were mapped to standard chemical identifiers such as PubChem, CAS, and HMDB using the metabolite ID conversion tool in MetaboAnalyst 5.0, and the data harmonization results were manually checked. A web application DeepVenn was then used for the comparison and visualization of two lists.

### Statistical analyses

All general analyses were performed using GraphPad Prism 9.3.1 unless specified. Data were presented as the number *n* and percent (%) for categorical variables and median (IQR) or mean ( $\pm$ standard deviation) for continuous variables. Differences between TD and risk groups were assessed by Student's *t*-test (parametric), Mann–Whitney *U* test (non-parametric) for continuous variables, and Fisher's exact test for categorical variables. A two-tailed *P* value < 0.05 was considered statistically significant. Sample size calculations for future validation studies were performed using Pearson's correlation-based effect size of 0.25, two-sided  $\alpha$  of 0.05, and  $\beta$  value of 0.8 (<https://sample-size.net/correlation-sample-size/>).

### Metabolome-exposome inter-omics correlation networks

Spearman's rank correlation was used to build the correlations between HM metabolome and exposome data using the processed *z*-scores. Spearman correlation *r* values, false discovery rate-adjusted *P* values, and *q* values were generated with all chemical pairs. Only correlations between each pair variable with absolute *r* > 0.5, and *q* < 0.05 were used to construct the final correlation networks by MetScape 3.0, a plug-in for Cytoscape 3.8.2.<sup>32</sup>

### Multi-analyte predictive model construction and evaluation

The predictive performance was evaluated using the area under the receiver operator characteristic (AUROC) curve and nomograms were also created using R 4.0.5 with corresponding R packages. Confidence intervals (CI) for the ROC curves were calculated by bootstrap resampling. The models were validated by permutation analysis (1000 times). Sensitivity, specificity, and accuracy were calculated by 2×2 contingency table

analyses. Matthews correlation coefficient (MCC) was also calculated, which provided complementary information to common binary classification metrics, especially for class-imbalanced datasets.<sup>33</sup> See details in the Supplementary Methods.

## RESULTS

### Participant and sample selection

Figure 1 illustrates the participant and sample selection. After a bioinformatic review of all frozen and archived samples contained in the human milk biobank at the University of California, San Diego, we identified 269 qualified milk samples that met the inclusion and exclusion criteria for this study. From these samples, we identified 23 samples from infants who were found to be at risk for NDD by ASQ2 testing between 12 and 25 months of age. We next matched TD control samples to the at risk (AR) samples by frozen storage time, infant sex, and age at the time of milk sample collection and selected the 60 best-matched samples as controls for metabolomic and exposomic analysis. One male control infant was excluded to avoid genetic ascertainment bias because his twin sister qualified as a match. Participant and sample characteristics are summarized in Supplementary Table S1.

### Study population

Our study cohort included 82 mother–infant pairs, out of which 40 infants were male, and 42 were female (Table 1). Mothers were primarily non-Hispanic (91.5%), Caucasian (85.4%) with an annual household income  $\geq$ US\$60,000 (80.5%), and already had a child (59.8%) (Table 1). At the time of milk sample collection, the mean maternal age was  $33.7 \pm 4.4$  years old, and the mean age of the infants was  $4.6 \pm 2.5$  months. The participants were periodically followed-up, and at the time of ASQ2 assessment (mean child age of 14.2 months), 23 children screened positive for NDD (the Risk group), while 59 were classified as TD. The time interval between milk collection and ASQ2 test was 9.5 (7.1–11.0) months (median and IQR). There was a significant difference in ASQ2 normalized scores between the risk and TD groups for the risk group,

**Table 1.** Participant and sample characteristics.

	At risk, mean ± SEM (range)	Controls, mean ± SEM (range)	P value
Subjects enrolled	23	59	–
Sex	11 males	29 males	–
	12 females	30 females	
Gestational age at delivery (weeks)	37.6 ± 0.6 (28–40)	38.6 ± 0.3 (28–41)	ns
Maternal age (years)	34.6 ± 1.1 (23–46)	33 ± 0.4 (23–39)	ns
Maternal BMI (kg/m <sup>2</sup> )	26.2 ± 1.2 (19.1–40.7)	25.7 ± 0.6 (18.8–42.0)	ns
Parity (percent >1 child)	35% (8 of 23)	49% (29 of 59)	ns <sup>a</sup>
Percent preterm infants (<36 weeks)	13% (3 of 23)	3.4% (2 of 59)	ns <sup>a</sup>
Percent C-section delivery	26% (6 of 23)	29% (17 of 59)	ns <sup>a</sup>
Percent singleton pregnancies	96% (22 of 23)	97% (57 of 59)	ns <sup>a</sup>
Birth weight (g)	2950 ± 174 (1040–4139)	3311 ± 72 (1110–4507)	ns <sup>b</sup>
1-min APGAR score	7.8 ± 0.4 (4–9)	7.8 ± 0.2 (3–9)	ns
5-min APGAR score	8.7 ± 0.2 (7–9)	8.9 ± 0.1 (7–10)	ns
Infant age at time of milk collection (months)	4.1 ± 0.6 (0.4–8.3)	4.8 ± 0.3 (0.7–8.9)	ns
Time of day of milk collection (24-h clock)	1248 ± 46 (0730–2300)	1306 ± 29 (0645–2400)	ns
Milk sample storage time at –80 °C (years)	2.8 ± 0.1 (2.2–3.8)	3.0 ± 0.1 (1.9–5.1)	ns
Age at ASQ2 testing (months)	14.2 ± 0.7 (7.4–25.4)	14.2 ± 0.4 (12–25.4)	ns
Interval between milk collection and ASQ2 testing (months)	10.2 ± 0.8 (4.0–24.4)	9.5 ± 0.5 (4.5–21.9)	ns
Percent with ≥2 ASQ2 tests administered before 25 months	4.3% (1 of 23)	6.8% (4 of 59)	ns <sup>a</sup>
Age-normalized ASQ2 communication (Z-score)	–1.9 ± 0.2 (–3.2 to –0.92)	0.18 ± 0.1 (–1.8 to +1.0)	0.0001
Annual household income			
<\$10,000	0	1	ns
\$10,001–\$59,999	5	7	ns
≥\$60,000	18	51	ns
Ethnicity			
White, non-Hispanic	16	48	ns
White, Hispanic	2	4	ns
Black, non-Hispanic	0	1	ns
Black, Hispanic	1	0	ns
Asian	3	5	ns
Native American	1	1	ns

ns not significant.

<sup>a</sup>Fisher's exact test.

<sup>b</sup>t-test with Welch's correction for unequal variances.

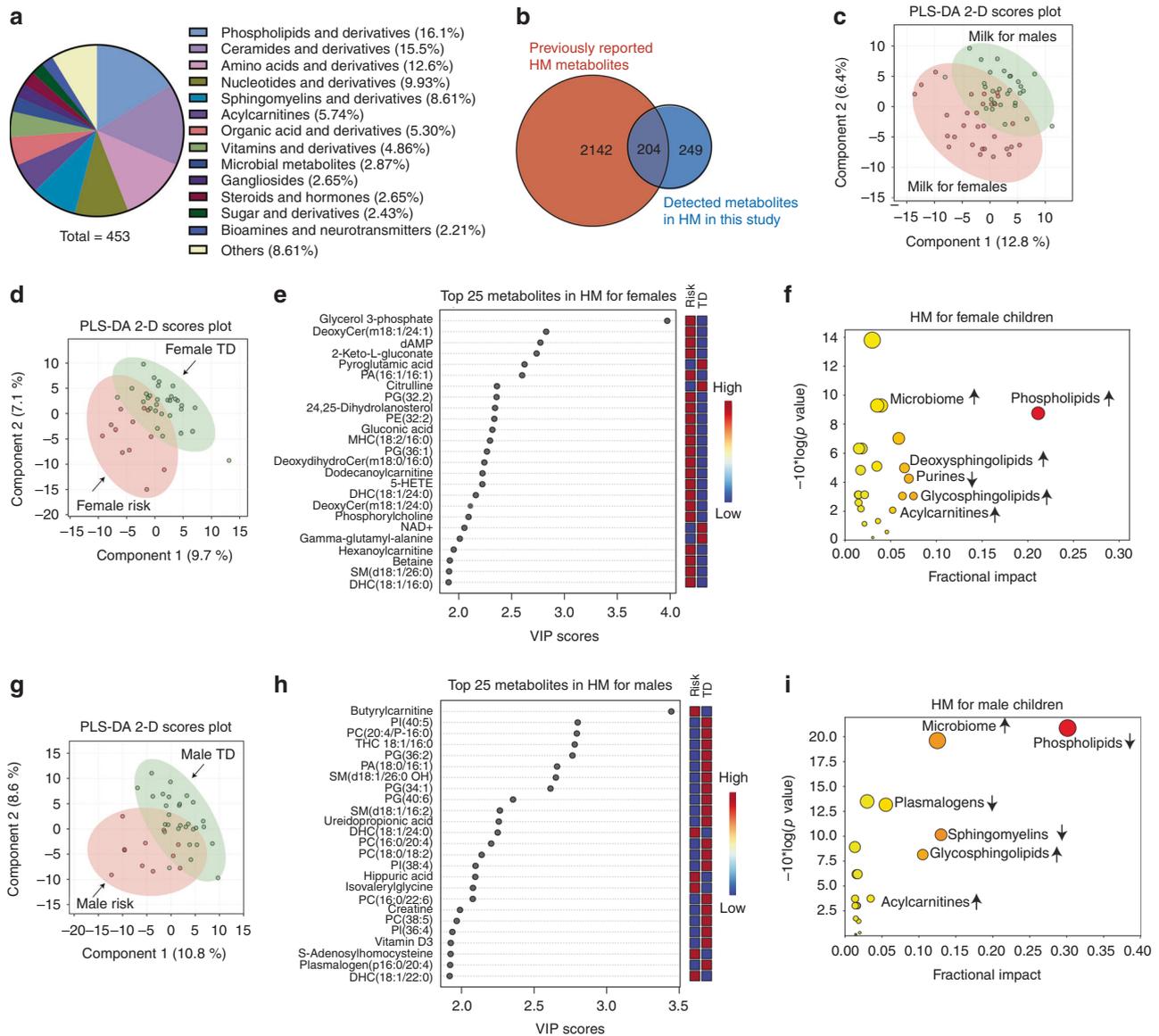
$P < 0.0001$  (Table 1). No significant differences were observed for other clinical variables between the risk and TD groups such as mother's age, body mass index, gestational age at delivery, and infant birth weight (Table 1).

### Sex-specific metabolomic features in human milk for infants with risk of future NDD

We performed the comprehensive characterization of the human milk metabolome using a multi-platform, targeted, broad-spectrum approach. The representative chromatograms for the metabolomic analysis of an HM sample are shown in Supplementary Fig. S1, and the raw data are reported in Supplementary Table S4. A total of 453 endogenous metabolites were detected in the HM samples with phospholipids as the most abundant chemical classes, followed by ceramides, amino acids, nucleotides, sphingomyelins, and acylcarnitines (Fig. 2a). Organic acids, vitamins, microbial metabolites, gangliosides, steroids, sugars, amines, and neurotransmitters were also detected. We compared these detected HM metabolites in our study with the existing HM metabolome data and found that 249 metabolites were reported for the first time in the HM samples in our study (Fig. 2b and

Supplementary Table S5). Furthermore, partial least square discriminant analysis (PLS-DA) revealed distinct differences in HM metabolome between male and female TD infants (Fig. 2c). HM for males had higher levels of virtually all the lipid classes we measured than those for the females, including phospholipids, sphingomyelins, cardiolipins, plasmalogens, and acylcarnitines (Supplementary Fig. S2a, b). We therefore analyzed male and female infants separately to identify the metabolic biosignatures in HM for infants with the risk of future NDD compared with TD infants.

In the PLS-DA 2D plot of the metabolomic results (Fig. 2d), the female risk children had strong early chemical signatures in their HM metabolome that were distinct from the TD controls. The HM metabolites that separate female risk from female TD children in the PLS-DA model were ranked by their variable importance in projection (VIP) scores in Supplementary Table S6, and the top 25 most discriminating metabolites are shown in Fig. 2e. Pathway impact analysis showed that lipid abnormalities constituted about 63% of the top 8 pathway changes in females (Supplementary Table S7). Compared with female TD children, the most striking change in lipid pathways was the increase of phospholipids,



**Fig. 2** Early biosignatures in human milk (HM) metabolome for children with the risk of future neurodevelopmental delay (NDD) compared with typically developing (TD) children. **a** Metabolites and their chemical classes detected in the HM samples in this study. **b** The comparison of HM metabolome coverage between previously reported milk metabolites and the metabolites detected in this study. **c** The HM metabolomic profiles differed substantially according to offspring sex ( $n = 29$  for male TD and 30 for female TD infants). **d** Partial least square discriminant analysis (PLS-DA) based on the HM metabolome data separates female infants with the future NDD risk ( $n = 12$ ) from female TD children ( $n = 30$ ). **e** The top 25 differential metabolites in the HM between the female risk and female TD groups in the PLS-DA model ranked by variable importance in projection (VIP) scores. **f** The metabolic pathways altered in the HM metabolome for female children in the risk group compared to those in the TD group. Increase (upper arrows) and decrease (down arrows) in the risk group. **g** PLS-DA reveals dramatically different metabolic phenotypes in the HM for male children between the risk ( $n = 11$ ) and TD ( $n = 29$ ) groups. **h** The top 25 differential metabolites in the HM for male infants between the risk and TD groups in the PLS-DA model ranked by their VIP scores. **i** Pathway impact analysis of altered biochemical pathways in HM for male children with higher risk of future NDD. Upper arrows: increase in the risk group. Down arrows: decrease in the risk group.

glycosphingolipids, deoxysphingolipids, and acylcarnitines in HM for female infants at risk for future NDD (Fig. 2f). The top non-lipid changes included the decrease of purines and the increase of microbiome metabolites (Fig. 2f). Similarly, PLS-DA showed distinct metabolomic differences in HM for the male infants between children with and without NDD risk (Fig. 2g). The discriminant HM metabolites between the male TD and male risk group in the PLS-DA model were ranked by their VIP scores in Fig. 2h and Supplementary Table S8. Males were also dominated by lipid abnormalities, which account for about 72% of the top 8 pathway changes (Supplementary Table S9 and Supplementary

Fig. S3). Glycosphingolipids and acylcarnitines were increased, while phospholipids, sphingomyelins, and plasmalogens were decreased in the HM for male risk infants (Fig. 2i). In the non-lipid pathways, most microbiome metabolites (6 out of 7) were increased in the male risk group (Fig. 2i and Supplementary Table S8).

Even though 13 biochemical pathways changes in HM metabolome were common to both male and female infants with the risk of future NDD, the direction of change was opposite for 7 pathways between sexes (Supplementary Fig. S4). For example, phospholipids and sphingomyelins were found to be increased in

HM for females at risk. In contrast, these metabolites were decreased in HM for males at risk. Nineteen metabolic pathways were altered in only one sex, among which deoxysphingolipids were increased in the HM for females at risk, and plasmalogens were decreased in the HM for males at risk (Supplementary Fig. S4).

### Human milk exposome and exposome-metabolome correlations

We next characterized the intrinsic human milk exposome to explore the xenobiotics and their metabolomic responses in HM associated with the risk of NDD in the offspring. Out of 937 common anthropogenic compounds screened (Supplementary Tables S2 and S3), 61 were detected in our HM samples and confirmed with the purified standards. These included flame retardants, persistent organic pollutants, phthalates, pesticides, personal care products, pharmaceuticals, food additives, and others (Fig. 3a and Supplementary Fig. S5). There were no significant sex differences in HM exposome between male and female infants (Fig. 3b). The median number of positive hits per sample was 20 (IQR: 20–24) in HM for TD infants and 21 (IQR: 20–23) for infants with future NDD risk ( $P=0.28$ ) (Fig. 3c). Interestingly, we found a significantly higher level of caffeine in HM for the children with NDD risk than that in the TD group ( $P=0.014$ , Fig. 3d).

We then explored the correlations between HM exposome and the metabolome to explore the metabolomic responses to external exposures. As shown in Fig. 3e and Supplementary Table S10, the exposome in the HM for female infants had significant positive correlations with the endogenous metabolites in the biochemical pathways of phospholipid, deoxysphingolipid, eicosanoid, acylcarnitines, and negative correlations with glycosphingolipids (Spearman's correlation  $r>0.5$  or  $<-0.5$  and  $q$  value  $<0.05$ ). Besides the positive correlations with phospholipid and deoxysphingolipid metabolism as observed in the female group, the HM exposome for male infants was positively correlated with ceramides and negatively correlated with sphingomyelins (Fig. 3f and Supplementary Table S11).

### Multi-analyte models for early prediction of the risk of neurodevelopmental delay

We used the combination of data-driven approaches using ML methods and knowledge-based methods (biochemical pathway analysis) to select the optimal features from HM metabolome and exposome for predicting the future risk of NDD in the offspring (Supplementary Fig. S6). Metabolomic and exposomic data are highly dimensional, and ML can reduce the dimensionality of omics data and improve the risk prediction accuracy.<sup>34</sup> Random forest (RF) analysis is a common ML method for omics-based predictive models due to its resilience to high dimensionality, insensitivity to noise, and robustness to overfitting.<sup>35</sup> We used the RF algorithm to identify which variables have a more determinant impact on the prediction outcomes. The importance of the variables in HM metabolome and exposome contributed to the classification between the risk and TD group was ranked by mean decrease accuracy scores in Supplementary Table S12 for female infants and Supplementary Table S13 for males. PLS-DA was used for feature reduction and kNN clustering analysis was performed to help reduce the feature redundancy (Supplementary Tables S14 and S15, Supplementary Figs. S7 and S8). Mann-Whitney  $U$  test was further applied to verify the results of ML. Finally, pathway enrichment analysis was conducted, and the optimal features were refined to cover different biochemical pathways (Supplementary Figs. S9 and S10).

Figure 4a shows the group differences of the selected analytes in the model for predicting NDD risk for female infants. DeoxyCer (m18:1/24:1), hexanoylcarnitine, PA (16:1/16:1), and caffeine were significantly higher in HM for female infants in the risk group than

those in the TD group and negatively correlated with ASQ2 scores (Fig. 4a and Supplementary Fig. S11a). In contrast, the NAD<sup>+</sup> level was significantly lower in the HM samples for the risk children compared with TDs and positively correlated with ASQ2 scores. ROC analysis yielded a AUROC of 0.81 (95% CI: 0.66–0.96, permutation  $P=0.003$ ) (Fig. 4b). In addition, the developed predictive model had the sensitivity of 83.3% (95% CI: 51.5–97.9), specificity of 76.7% (95% CI: 57.7–90.1), accuracy of 78.6% (95% CI: 63.2–89.7), and MCC of 0.55 for predicting future NDD risk for female infants (Table 2). A nomogram was then developed to provide a quantitative method for predicting the individual risk of NDD in females (Fig. 4c). The nomogram maps the predicted risk probabilities into element points on a scale from 0 to 100 in a user-friendly graphical interface. The element point for each metabolite in the model is calculated based on its Z-score. For example, an HM sample with DeoxyCer (m18:1/24:1) of  $-1$  (Z-score) will be assigned to 30 element points (the upper scale in Fig. 4c). The total points accumulated by all covariates in the nomogram correspond to the predicted probability of NDD risk for a child.

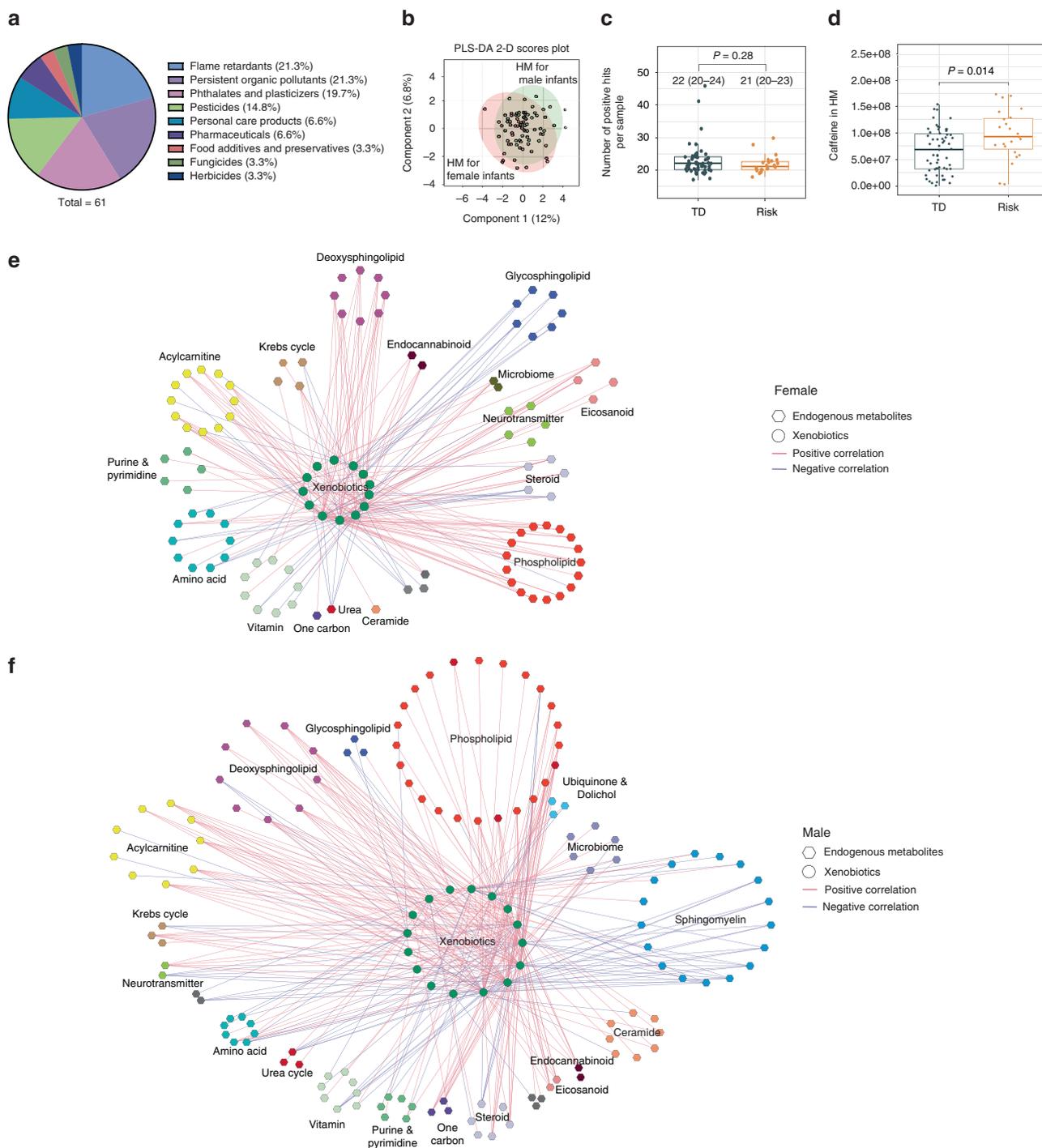
A risk prediction model for males was also developed using the selected optimal features (Fig. 4d). Butyrylcarnitine, DHC (18:1/24:0), and caffeine were significantly higher in HM for female infants in the risk group than those in the TD group and negatively correlated with ASQ2 scores (Fig. 4d and Supplementary Fig. S11b). In contrast, the levels of plasmalogen (p16:0/20:4), and NAD<sup>+</sup> were significantly lower in the HM samples for the male risk children compared with TDs and positively correlated with ASQ2 scores. The predictive performance evaluated using ROC analysis showed an AUROC of 0.79 (95% CI: 0.61–0.93, permutation  $P=0.012$ ) (Fig. 4e). Other performance metrics for the predictive model for males are listed in Table 2 and the nomogram is shown in Fig. 4f.

### Sample size estimation for validation study

We estimated the sample size required for a future large validation study by Pearson's correlation-based effect size (variance explained)  $r$  of 0.25. In the current discovery phase study, 22 metabolites in the female group, and 34 metabolites in the male group had absolute  $r$  values  $>0.25$ , which represents a medium to large effect size. Based on these calculations, a total of 124 females and 124 males, with 62 children at risk and 62 TD controls from each sex, will be required for a future validation study.

### DISCUSSION

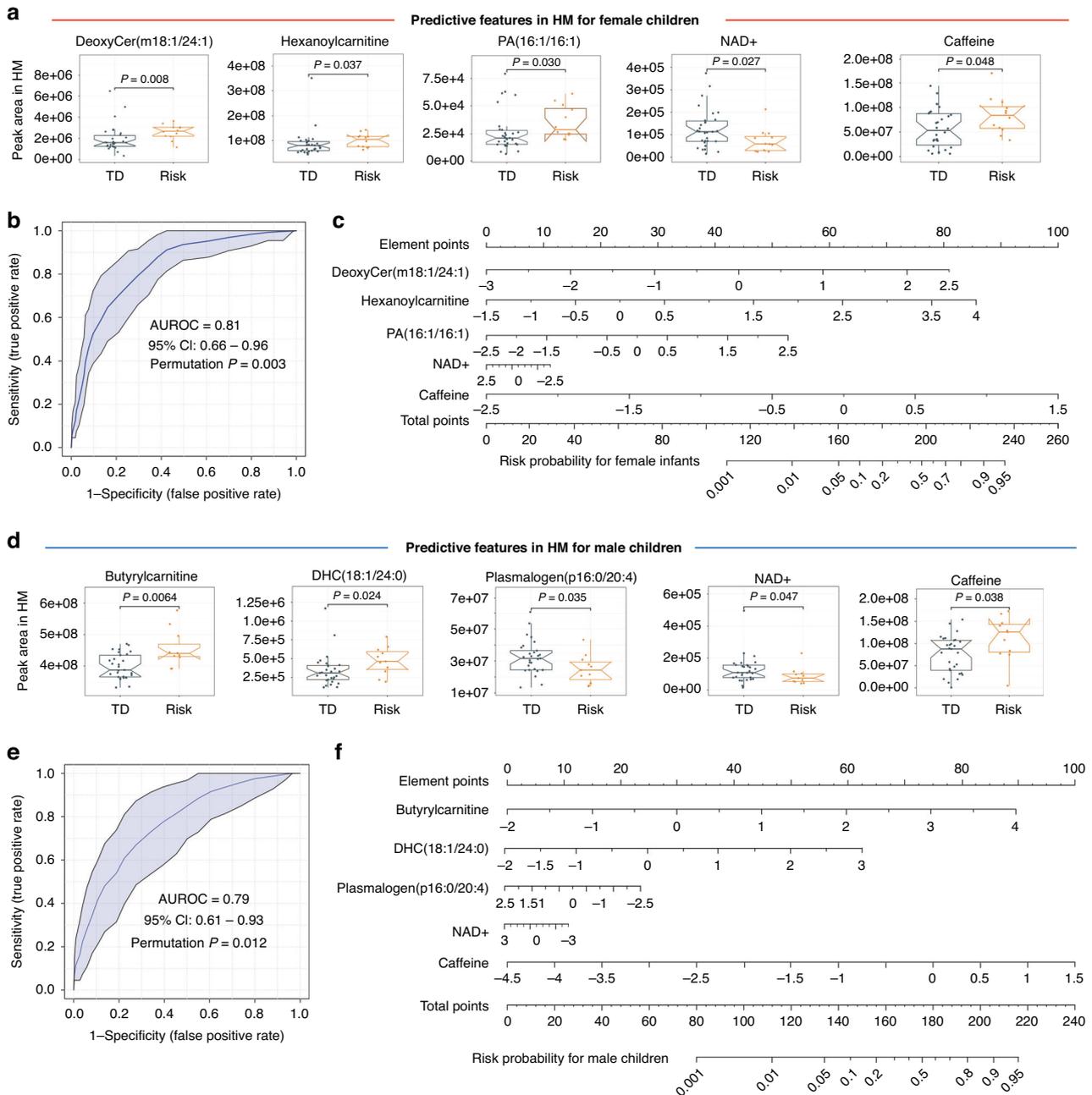
To our knowledge, this is the first metabolomics and exposomics study to demonstrate an early signature in human milk for infants at risk of future NDD. The primary patterns of metabolic alterations we found include critical metabolites involved in myelination and brain development, such as deoxysphingolipids, acylcarnitines, sphingomyelins, phospholipids, and glycosphingolipids, including several globosides and cerebroside. Predicting cognitive outcomes based on infant assessments is known to be challenging because problems such as language impairment, learning disabilities, and social-emotional disorders emerge over time and cannot be evaluated accurately at an early age.<sup>3</sup> In this study, by the combination of ML methods and knowledge-based approaches, we successfully developed sex-specific, multi-analyte predictive models from the HM metabolome and exposome to predict future risk of NDD in the toddler age with over 75% accuracy. Our models include novel metabolic and exposomic features in HM that are likely to be informative of disease etiology. If replicated, this finding could have a significant impact on clinical practice by permitting infants to be stratified according to their future risk and new preventive treatments to be tested in clinical trials.



**Fig. 3 Human milk (HM) exposome and metabolome  $\times$  exposome correlations.** **a** A pie chart shows the chemical class breakdown of the 61 detected xenobiotics out of 937 anthropogenic compounds screened in the HM samples in this study. **b** Partial least square discriminant analysis (PLS-DA) demonstrates the large overlap of HM exposome profiles between male ( $n = 40$ ) and female infants ( $n = 42$ ). **c** The number of positive hits per HM sample was not significantly different between children with the risk of neurodevelopmental delay at the median age of 13.5 months (risk group,  $n = 23$ ) and typically developing children (TD group,  $n = 59$ ). **d** The concentration of caffeine in the HM of the risk group ( $n = 23$ ) was significantly higher than that in the TD group ( $n = 59$ ). Mann–Whitney  $U$  test was conducted. **e, f** The metabolome  $\times$  exposome correlation networks in milk for female infants (**e**,  $n = 30$ ) and male infants (**f**,  $n = 29$ ). Only correlations with Spearman’s rank correlation  $r$  values  $> 0.5$  or  $< -0.5$  and  $q$  values  $< 0.05$  are shown in the networks.

Offspring sex-specific differences in human milk composition have been observed for energy content,<sup>36</sup> total protein,<sup>37</sup> fat content,<sup>38</sup> and microbiota diversity.<sup>39</sup> Beyond the existing literature, in this study, we show dramatic differences in HM metabolome between male and female offspring. HM for male

offspring had higher phospholipids, sphingomyelins, cardiolipins, acylcarnitines, and plasmalogens than those in the HM produced for female infants. Moreover, we found that the HM metabolome responds to the infant’s stress differentially between the sexes. Compared to the female TD children, participants with female



**Fig. 4** Multi-analyte models based on the human milk (HM) metabolome and exposome predict the children's future neurodevelopmental delay risk. **a** The selected analytes in the predictive model for female infants. The notched box plots represent the minimum and maximum values (whiskers), the IQR (length of the box), and the 95% CI around the median (notches).  $n = 12$  for the risk group and  $n = 30$  for the TD group. **b** The ROC curve for female infants. **c** The nomogram for predicting the risk of NDD in females using the selected analytes from HM metabolome and exposome. Z-scores were used. **d** The selected predictive analytes for male children.  $n = 29$  for the male TD group and 11 for the male risk group. **e** The ROC curve for male children. **f** The nomogram for predicting the risk of future NDD in male infants using the Z-scores of the selected analytes.

infants at risk produced more phospholipids in their HM to meet their children's needs. In contrast, phospholipids in HM for male risk infants were lower than those in the male TD group. Even though the HM exposome did not differ substantially according to the infant sex, the inter-omics analysis revealed significant positive correlations between the xenobiotics and ceramides in HM only for male children.

Low levels of certain critical metabolites in human milk may logically have important effects on infant brain development. In this study, plasmalogens, and NAD<sup>+</sup> were observed to be lower in

the HM for male infants in the risk group compared to the TD group. Plasmalogens are the main ethanolamine phospholipids in brain myelin and neural membranes, and because of their unique molecular structure, plasmalogens facilitate rapid membrane fusion, which is critical for synaptic neurotransmission.<sup>40</sup> Our metabolomic analysis confirms the presence of plasmalogens from the phosphatidylethanolamine family in HM, which have been implicated in infant brain development.<sup>41</sup> Increased HM levels of plasmalogens were reported to be associated with a remarkable increase in brain size in preterm infants.<sup>42,43</sup> NAD<sup>+</sup> is a

**Table 2.** Performance analysis of the multi-analyte discriminator in the human milk metabolome and exposome for predicting future neurodevelopmental delay.

Multi-analyte predictive model	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Accuracy (%) (95% CI)	Matthews correlation coefficient (MCC)
Female infants	83.3 (51.6–97.9)	76.7 (57.7–90.1)	78.6 (63.2–89.7)	0.55
Male infants	72.7 (39.1–93.9)	79.3 (60.3–92.1)	77.5 (61.5–89.2)	0.49

$N = 30$  for female TD infants and 12 female risk infants, and  $N = 29$  male TD infants and 11 male risk infants. Values were calculated based on  $2 \times 2$  contingency table analyses. The predictive model for female infants contains five metabolites from the human milk (HM), including DeoxyCer (m18:1/24:1), Hexanoylcarnitine, PA (16:1/16:1), NAD<sup>+</sup>, and caffeine. The predictive model for male infants contains six analytes: Butyrylcarnitine, SM (d18:1/26:0 OH), DHC (18:1/24:1), plasmalogen (p16:0/20:4), NAD<sup>+</sup>, and caffeine.

central redox cofactor of virtually all metabolic processes. Postpartum supplementation of nicotinamide riboside to rodent mothers increased NAD<sup>+</sup> content in the rodent milk, leading to improved neurobehavioral development of the offspring, whose advantages persist into adulthood.<sup>44</sup>

Compared to the TD group, we also observed higher levels of some essential metabolites for brain development in mother's milk for the infants with future NDD risk in a sex-specific manner, including higher phospholipids and sphingomyelins for female risk infants, and higher glycosphingolipids and acylcarnitines for infants in the risk group of both sexes. Interestingly, increased glycosphingolipids, phospholipids, and acylcarnitines are typical metabolic features of cell danger response (CDR).<sup>45</sup> During the lactation period, the mother and her breastfed offspring maintain a reciprocal regulation of responsiveness to stress in both rodents and human beings.<sup>46</sup> Activation of CDR represents the mother's metabolic responses to the offspring's stress, resulting in the tailoring of breast milk composition to maximize health protection and development for infants at risk.

Caffeine is a methylxanthine that acts as a central nervous system stimulant and naturally occurs in several foods. Neurostimulation occurs because caffeine is an inhibitor of the widely expressed purinergic adenosine receptors 1 and 3 (P1A1 and P1A3, also abbreviated as ADORA1 and ADORA3 receptors). Here, we show that HM for the infants in the risk group had higher caffeine than that for the TD infants. The effects of maternal caffeine consumption on breastfed children are inconclusive, and both negative and positive outcomes have been reported in the literature.<sup>47</sup> The potential effects of early caffeine exposure on the long-term neurodevelopment of breastfed infants require further investigation.

A limitation of the current study was the relatively small size of the cohort of children at risk. Sample size calculations estimated that at least 124 males and 124 females, with 62 TD and 62 AR of each sex, will be needed for a larger validation phase study. It should also be noted that the nutrient content of HM varies in samples collected at different times of the day.<sup>16,17</sup> The samples in this study were well matched with a mean sampling time around 1 pm. Another limitation was that the outcome was measured by maternal report using the ASQ at about one year of age, and thus, the assessment of risk for NDD was not diagnostic. However, the ASQ screening instrument is widely used in pediatric practice and can be an important early indicator of future risk, leading to potential further evaluation, ongoing monitoring, or specialized intervention.

## CONCLUSION

In summary, multi-omics profiling of the maternal milk metabolome and exposome revealed significant differences between children at risk for neurodevelopmental delay (NDD) and typically developing (TD) infants. Other well-known risk factors such as prematurity, low birth weight, perinatal difficulties, and socio-economic variables were well controlled and were unable to

account for the differences between groups in this study. The early biosignature in human milk for infants who later screened positive for NDD was dominated by the sex-specific abnormalities in deoxysphingolipids, sphingomyelins, plasmalogens, phospholipids, and acylcarnitines. Multi-analyte models of the biosignature of milk samples collected before 9 months of age showed over 75% accuracy in predicting NDD documented nearly one year later, raising the possibility that milk metabolomics might provide an early warning of later developmental risk. Caution is warranted in interpreting these results. This was a small pilot study. Larger validation studies will be needed to determine the sensitivity, specificity, cost/benefit, and clinical practicality of milk metabolomics as a tool in pediatrics. However, if validated in larger studies, milk metabolomics might one day be used as another early screening tool to help identify infants at risk of NDD with the aim of instituting treatment before the first symptoms appear.

## DATA AVAILABILITY

The raw metabolome and exposome data for this study are provided in Supplementary Tables S3 and S4. In-house python tools used in this study can be found on GitHub (<https://github.com/BDNav/PeakVetting>).

## REFERENCES

- Villagomez, A. N. et al. Neurodevelopmental delay: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. *Vaccine* **37**, 7623–7641 (2019).
- Zablotsky, B. et al. Prevalence and trends of developmental disabilities among children in the United States: 2009–2017. *Pediatrics* **144**, e20190811 (2019).
- Glascoe, F. P., Gellasch, P. & Chen, V. When do clinicians decide to screen children for mental health-behavioral-developmental delays/disorders: is it time to reconsider policy recommendations? *J. Pediatr.* **206**, 248–255 (2019).
- Bar, S., Milanaik, R. & Adesman, A. Long-term neurodevelopmental benefits of breastfeeding. *Curr. Opin. Pediatr.* **28**, 559–566 (2016).
- Boucher, O. et al. Association between breastfeeding duration and cognitive development, autistic traits and ADHD symptoms: a multicenter study in Spain. *Pediatr. Res.* **81**, 434–442 (2017).
- Belfort, M. B. et al. Infant feeding and childhood cognition at ages 3 and 7 years: effects of breastfeeding duration and exclusivity. *JAMA Pediatr.* **167**, 836–844 (2013).
- Lyons, K. E., Ryan, C. A., Dempsey, E. M., Ross, R. P. & Stanton, C. Breast milk, a source of beneficial microbes and associated benefits for infant health. *Nutrients* **12**, 1039 (2020).
- Garwolinska, D., Namiesnik, J., Kot-Wasik, A. & Hewelt-Belka, W. Chemistry of human breast milk—a comprehensive review of the composition and role of milk metabolites in child development. *J. Agric. Food Chem.* **66**, 11881–11896 (2018).
- Boquien, C. Y. et al. Breast milk protein content at week 3 after birth and neurodevelopmental outcome in preterm infants fed fortified breast milk. *Eur. J. Nutr.* **60**, 3959–3969 (2021).
- Gay, M. C. L. et al. Worldwide variation in human milk metabolome: Indicators of breast physiology and maternal lifestyle? *Nutrients* **10**, 1151 (2018).
- Ojo-Okunola, A., Cacciatore, S., Nicol, M. P. & du Toit, E. The determinants of the human milk metabolome and its role in infant health. *Metabolites* **10**, 77 (2020).
- Cerdo, T., Dieguez, E. & Campoy, C. Infant growth, neurodevelopment and gut microbiota during infancy: which nutrients are crucial? *Curr. Opin. Clin. Nutr. Metab. Care* **22**, 434–441 (2019).

13. Eriksen, K. G., Christensen, S. H., Lind, M. V. & Michaelsen, K. F. Human milk composition and infant growth. *Curr. Opin. Clin. Nutr. Metab. Care* **21**, 200–206 (2018).
14. Ottolini, K. M., Andescavage, N., Keller, S. & Limperopoulos, C. Nutrition and the developing brain: the road to optimizing early neurodevelopment: a systematic review. *Pediatr. Res.* **87**, 194–201 (2020).
15. Zielinska, M. A., Hamulka, J., Grabowicz-Chadzynska, I., Brys, J. & Wesolowska, A. Association between breastmilk LC PUFA, carotenoids and psychomotor development of exclusively breastfed infants. *Int. J. Environ. Res. Public Health* **16**, 1144 (2019).
16. Hahn-Holbrook, J., Saxbe, D., Bixby, C., Steele, C. & Glynn, L. Human milk as “Chrononutrition”: implications for child health and development. *Pediatr. Res.* **85**, 936–942 (2019).
17. Ruel, M. T., Dewey, K. G., Martinez, C., Flores, R. & Brown, K. H. Validation of single daytime samples of human milk to estimate the 24-h concentration of lipids in urban Guatemalan mothers. *Am. J. Clin. Nutr.* **65**, 439–444 (1997).
18. Andreas, N. J., Kampmann, B. & Mehring Le-Doare, K. Human breast milk: a review on its composition and bioactivity. *Early Hum. Dev.* **91**, 629–635 (2015).
19. Karthikeyan, B. S., Ravichandran, J., Aparna, S. R. & Samal, A. Exhumid: a curated resource and analysis of exposome of human milk across India. *Chemosphere* **271**, 129583 (2021).
20. Reh, R. K. et al. Critical period regulation across multiple timescales. *Proc. Natl Acad. Sci. USA* **117**, 23242–23251 (2020).
21. LaKind, J. S. et al. Infant dietary exposures to environmental chemicals and infant/child health: a critical assessment of the literature. *Environ. Health Perspect.* **126**, 96002 (2018).
22. Tai, P. T. et al. Dioxin exposure in breast milk and infant neurodevelopment in Vietnam. *Occup. Environ. Med.* **70**, 656–662 (2013).
23. Pham, N. T. et al. Perinatal dioxin exposure and neurodevelopment of 2-year-old Vietnamese children in the most contaminated area from agent orange in Vietnam. *Sci. Total Environ.* **678**, 217–226 (2019).
24. Kim, S. et al. Association between maternal exposure to major phthalates, heavy metals, and persistent organic pollutants, and the neurodevelopmental performances of their children at 1 to 2 years of age—CHECK cohort study. *Sci. Total Environ.* **624**, 377–384 (2018).
25. Chao, H. R., Tsou, T. C., Huang, H. L. & Chang-Chien, G. P. Levels of breast milk PBDEs from southern Taiwan and their potential impact on neurodevelopment. *Pediatr. Res.* **70**, 596–600 (2011).
26. Bandoli, G., Bertrand, K., Saor, M. & Chambers, C. D. The design and mechanics of an accessible human milk research biorepository. *Breastfeed. Med.* **15**, 155–162 (2020).
27. Squires, J., Potter, L. & Bricker, D. *The ASQ User's Guide for the Ages & Stages Questionnaires: A Parent-completed, Child-monitoring System* (Paul H Brookes Publishing, Baltimore, 1995).
28. Bayley, N. & Infant, S. *Bayley Scales of Infant and Toddler Development 3rd edn* (Harcourt Assessment, San Antonio, 2006).
29. Li, K., Naviaux, J. C., Bright, A. T., Wang, L. & Naviaux, R. K. A robust, single-injection method for targeted, broad-spectrum plasma metabolomics. *Metabolomics* **13**, 122 (2017).
30. Li, K., Naviaux, J. C., Monk, J. M., Wang, L. & Naviaux, R. K. Improved dried blood spot-based metabolomics: a targeted, broad-spectrum, single-injection method. *Metabolites* **10**, 82 (2020).
31. Liu, Y., Liang, Y. & Wishart, D. Polysearch2: a significantly improved text-mining system for discovering associations between human diseases, genes, drugs, metabolites, toxins and more. *Nucleic Acids Res.* **43**, W535–W542 (2015).
32. Basu, S. et al. Sparse Network modeling and metscape-based visualization methods for the analysis of large-scale metabolomics data. *Bioinformatics* **33**, 1545–1553 (2017).
33. Chicco, D. & Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 6 (2020).
34. Huang, L. et al. Machine learning of serum metabolic patterns encodes early-stage lung adenocarcinoma. *Nat. Commun.* **11**, 3556 (2020).
35. Sarica, A., Cerasa, A. & Quattrone, A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. *Front Aging Neurosci.* **9**, 329 (2017).
36. Powe, C. E., Knott, C. D. & Konkin-Brittain, N. Infant sex predicts breast milk energy content. *Am. J. Hum. Biol.* **22**, 50–54 (2010).
37. Galante, L. et al. Sex-specific human milk composition: the role of infant sex in determining early life nutrition. *Nutrients* **10**, 1194 (2018).
38. Hosseini, M., Valizadeh, E., Hosseini, N., Khatibshahidi, S. & Raeisi, S. The role of infant sex on human milk composition. *Breastfeed. Med.* **15**, 341–346 (2020).
39. Moossavi, S. et al. Composition and variation of the human milk microbiota are influenced by maternal and early-life factors. *Cell Host Microbe* **25**, 324–335.e324 (2019).
40. Dorninger, F. et al. Homeostasis of phospholipids—the level of phosphatidylethanolamine tightly adapts to changes in ethanolamine plasmalogens. *Biochim. Biophys. Acta* **1851**, 117–128 (2015).
41. Moukartzel, S., Dyer, R. A., Keller, B. O., Elango, R. & Innis, S. M. Human milk plasmalogens are highly enriched in long-chain PUFAs. *J. Nutr.* **146**, 2412–2417 (2016).
42. Dekaban, A. S. Changes in brain weights during the span of human life: Relation of brain weights to body heights and body weights. *Ann. Neurol.* **4**, 345–356 (1978).
43. Alexandre-Gouabau, M. C. et al. Breast milk lipidome is associated with early growth trajectory in preterm infants. *Nutrients* **10**, 528 (2018).
44. Ear, P. H. et al. Maternal nicotinamide riboside enhances postpartum weight loss, juvenile offspring development, and neurogenesis of adult offspring. *Cell Rep.* **26**, 969–983 e964 (2019).
45. Naviaux, R. K. Metabolic features of the cell danger response. *Mitochondrion* **16**, 7–17 (2014).
46. Walker, C. D. et al. Mother to infant or infant to mother? Reciprocal regulation of responsiveness to stress in rodents and the implications for humans. *J. Psychiatry Neurosci.* **29**, 364–382 (2004).
47. McCreedy, A., Bird, S., Brown, L. J., Shaw-Stewart, J. & Chen, Y. F. Effects of maternal caffeine consumption on the breastfed child: a systematic review. *Swiss Med. Wkly.* **148**, w14665 (2018).

## AUTHOR CONTRIBUTIONS

Conceptualization: C.C., K.L., K.B., R.K.N. IRB approvals: C.C., R.K.N. Participant enrollment: K.B., C.C. Metabolomics and exposomics: K.L., J.C.N., S.S.L., L.W., R.K.N. Data analysis and visualization: J.M.M., K.L., K.B., A.W., R.K.N., C.C. Project supervision: C.C., R.K.N. Funding: C.C., R.K.N. Writing original draft: K.L., K.B., C.C., R.K.N. Writing review and editing: all coauthors.

## FUNDING

This project was supported in part by gifts from the UCSD Christini Fund, the Lennox Foundation, the JMS Fund, the Malone Family Foundation, the Westreich Foundation, the Daniel and Kelly White Family, the UCSD Mitochondrial Research Fund, and grassroots support from over 2000 individuals from around the world who have each provided gifts in the past year to support Naviaux Lab research. Funding for the mass spectrometers was provided by a gift from the Jane Botsford Johnson Foundation. Funding for human milk sample processing and storage was provided by the UCSD Altman Clinical Translational Research Institute (ACTRI) funded by the National Institutes of Health (NIH) National Center for Advancing Translational Sciences (NCATS) under Award Number UL1TR001442. The funders had no role in the study design, data collection and analysis, decision to publish, where to publish, or preparation of the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All the study protocols were approved by the Institutional Review Board (IRB) at UC San Diego, and written consent was obtained upon the collection of the HM samples under C.C.'s UC San Diego protocol #130658. Metabolomics and exposomics were conducted on deidentified samples under R.K.N.'s UC San Diego IRB-approved protocol #140072.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41390-022-02283-6>.

**Correspondence** and requests for materials should be addressed to Robert K. Naviaux or Christina Chambers.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



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

© The Author(s) 2022