Prenatal exposure to perfluoroalkyl substances modulates neonatal serum phospholipids, increasing risk of type 1 diabetes

Aidan McGlinchey[1], Tim Sinioja[2], Santosh Lamichhane[3], Johanna Bodin[4], Heli Siljander[5], Alex M. Dickens[3], Dawei Geng[2], Cecilia Carlsson[2], Daniel Duberg[2], Jorma Ilonen[6,7], Suvi M. Virtanen[8,9], Hubert Dirven[4], Hanne Friis Berntsen[10,11], Karin Zimmer[10], Unni C. Nygaard[4], Matej Orešič[1,3,*], Mikael Knip[5,*], Tuulia Hyötyläinen[2,*]

[1]School of Medical Sciences, Örebro University, 702 81 Örebro, Sweden

[2]School of Science and Technology, Örebro University, 702 81 Örebro, Sweden

[3]Turku Bioscience Centre, University of Turku and Åbo Akademi University, 20520 Turku, Finland

[4]Division of Infection Control and Environmental Health, Norwegian Institute of Public Health, 0456 Oslo, Norway

[5]Children's Hospital, University of Helsinki and Helsinki University Hospital, 00290 Helsinki, Finland; Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, 00290 Helsinki, Finland

[6]Immunogenetics Laboratory, Institute of Biomedicine, University of Turku, 20014 Turku, Finland

[7]Department of Clinical Microbiology, Turku University Hospital, 20014 Turku, Finland

[8]Public Health Promotion Unit, National Institute for Health and Welfare, 00271 Helsinki, Finland

[9]Center for Child Health Research, Tampere University Hospital, 33520 Tampere, Finland

[10]Norwegian University of Life Sciences, 0102 Oslo, Norway

[11]National Institute of Occupational Health, 0363 Oslo, Norway

*Shared senior authorship.

Correspondence:

Matej Orešič; Email: matej.oresic@utu.fi

Mikael Knip; Email: mikael.knip@helsinki.fi

Tuulia Hyötyläinen; MTM Research Centre, School of Science and Technology, Örebro University, 702 81 Örebro, Sweden. Email: tuulia.hyotylainen@oru.se; Phone: +46 19 303487

Abstract

In the last decade, increasing incidence of type 1 diabetes (T1D) stabilized in Finland, a phenomenon that coincides with tighter regulation of perfluoroalkyl substances (PFAS). Here, we quantified PFAS to examine their effects, during pregnancy, on lipid-related and immune markers of T1D risk in children. In a mother-infant cohort (264 dyads), high PFAS exposure during pregnancy associated with decreased cord serum phospholipids and progression to T1D-associated islet autoantibodies in the offspring. This PFAS-lipid association was exacerbated by increased human leukocyte antigen-conferred risk of T1D in infants. Exposure to a single PFAS compound or a mixture of organic pollutants in non-obese diabetic mice resulted in a profile characterized by a similar decrease in phospholipids, a marked increase of lithocholic acid, and accelerated insulitis. Our findings suggest that PFAS exposure during pregnancy contributes to risk and pathogenesis of T1D in offspring.

INTRODUCTION

T1D is an autoimmune disease caused by destruction of insulin-secreting pancreatic beta-cells[1]. The strongest genetic risk factors for T1D are found within the human leukocyte antigen (HLA) gene complex[2], yet only 3-10% of individuals carrying HLA-conferred disease susceptibility develop T1D[3]. The role of environmental factors in T1D pathogenesis is thus obvious[4]. We and others previously observed that children progressing to T1D-associated islet autoantibody positivity or to overt T1D later in life have a distinct lipidomic profile characterized by decreased blood phospholipid levels, including sphingomyelins (SMs), within the first months of life, preceding the onset of islet autoimmunity[5, 6, 7] and even as early as at birth[8, 9]. The cause of these metabolic changes is currently poorly understood. The gut microbiome is known to affect host lipid metabolism[10] and is associated with progression to T1D[11, 12], particularly in the period after islet autoantibody seroconversion, but current data does not offer an explanation for the earlier changes in phospholipid levels[11].

The incidence of T1D has been increasing over the last decades in many industrialized countries[13]. However, for unknown reasons, this has stabilized in the last decade, particularly in the Nordic countries[14, 15]. Environmental triggers and specific co-morbidities are often implicated in T1D, such as enterovirus infection, diet, and obesity[4], yet these do not explain this observation. Obesity, for example, has not shown a concomitant decrease since 2005[16], and the number of severe enterovirus infections in Finland 2006-2010 increased, in fact, by 10 fold[15].

Notably, the time trend of human exposure levels to two widely-used industrial chemicals, namely, perfluorooctane sulfonate and perfluorooctanoic acid (PFOS and PFOA), does coincide with this trend in T1D incidence rate[15]. These two compounds belong to the group of per- and poly-fluoroalkyl substances (PFAS) which are widely-used in food packaging materials, paper and textile coatings, and fire-fighting foams[17]. The use of PFOS and PFOA has increased substantially since production started in the 1950s until the main, global manufacturer phased out its production of PFOS, PFOS-related substances and PFOA between 2000-2002. In the European Union, all uses of PFOS are now prohibited under Directive (2006/122/EC) which came into force in 2008 due to concerns regarding persistent effects in the environment and both bioaccumulation and toxic effects in humans. PFOA is still manufactured and a large number of other PFAS compounds are currently in use. With a half-life of up to five years for PFOS and two to four years for PFOA in humans, concentrations of PFOS and PFOA started to decrease in man only after *ca*. 2005, with the levels of many other PFAS still showing increasing trends[18, 19]. The main sources of exposure to PFAS in the general population are food and drinking water, with lesser sources including house dust and air. PFAS are transferred from mother to fetus *via* the placenta and to breast-fed infants *via* maternal milk[20].

Structurally, most PFAS resemble endogenous fatty acids, with fluorine substituted in place of hydrogen. Functionally, PFAS share some common features with bile acids, which are key metabolites

involved in the digestion and absorption of lipids in the small intestine as well as in the maintenance of lipid and glucose homeostasis[21]. Bile acids are excreted into the intestine and reabsorbed, and similar enterohepatic circulation has been suggested for PFOS and PFOA[22, 23]. It has been estimated that over 90% of PFOS and PFOA have to be reabsorbed in order to explain the long half-life of these compounds in humans[24, 25]. Bile acids can therefore potentially act as mediators, linking PFAS exposures and altered lipid metabolism. In fact, prenatal exposure to PFAS was recently shown to associate with worsening metabolic outcomes in the offspring, including impaired glucose tolerance[26].

There is a dearth of knowledge regarding PFAS as possible contributors to T1D pathogenesis, although a contribution to the development of T1D has been proposed *via* impaired beta/immune-cell functions and immunomodulation[27]. It has also been reported that PFOA and PFOS disrupt generation of human pancreatic progenitor cells[28]. Prenatal and early-life exposure to perfluoroundecanoic acid (PFUnDA) aggravated insulitis development in NOD mice[29]. Recently, elevated levels of PFOS were reported in children at the point of diagnosis of T1D[30].

Here we hypothesized that PFAS exposure *in utero* affects the phospholipid profile of newborn infants, which may contribute to increased T1D risk. In a mother-infant cohort study, we (1) analyzed metabolite profiles of pregnant mothers and their offspring at birth, (2) quantified selected PFAS in maternal samples during pregnancy, and (3) examined prenatal PFAS exposures in relation to neonatal metabolite profiles and progression to T1D-associated islet autoantibody positivity (AAb+) during follow-up. We then further experimentally examined the impact of PFAS exposure on lipidomic and bile acid profiles and the development of insulitis / autoimmune diabetes in NOD mice as well as verified our key findings in a prospective birth cohort study comprising children at risk for T1D.

RESULTS

Metabolomic analyses of the mother-infant cohort

A total of 264 mother-infant dyads were included in the study (**Figure 1**; **Supplementary Table 1**). Age at delivery was between 18.5 and 45.8 years, pre-pregnancy body mass index (BMI) was between 16.9 and 45.7 kg/m^2, with 62% of the mothers being normal weight (BMI 18.5-25). All babies were born after gestational week 35. 74 children at-risk for T1D were assayed for AAb+ during follow-up, and ten among these progressed to at least one islet autoantibody.

Serum concentrations of 25 PFAS compounds were determined in the mothers during the pregnancy, collected at two time points, one during pregnancy and one at delivery (**Supplementary Table 2**). The two most abundant PFAS were PFOS and PFOA, detected in all subjects. Our detected levels of PFOS and PFOA were lower than reported in previous studies[19], most of which used samples collected before 2010, and therefore before recent, noted decreases in population blood levels of PFOS and PFOA[31].

Metabolomic analyses were performed using two analytical platforms. Serum molecular lipids and polar metabolites were quantified from the mothers during pregnancy and from newborn infants (cord serum). Identified lipids (n = 206) and quantified polar metabolites (n = 35) were included in the final datasets. To reduce dimensional complexity and facilitate identification of global associations between metabolic profiles and maternal PFAS exposure, we first clustered the metabolites from all datasets into cluster variables using model-based clustering[32], followed by partial-correlation network analysis[33]. The optimum number of clusters for each dataset, as assessed by Bayesian Information Criterion (**Supplementary Figure 1**), returned ten Maternal Lipid Clusters (MLCs) and four Maternal polar Metabolite Clusters (MMCs) (**Supplementary Table 3**), while the cord serum lipidomics data yielded eight Child Lipid Clusters (CLCs) and four Child polar Metabolite Clusters (CMCs) (**Supplementary Table 4**).

Metabolic profiles in mothers associate with PFAS exposure

PFAS exposure impacted the maternal metabolome (**Figure 2**, **Supplementary Figure 2**), both at the cluster variable level as well as at the individual metabolite level. Total PFAS, as well as several individual PFAS, levels were positively associated with MMC1 (amino acids, saturated free fatty acids and cholesterol). At the individual metabolite level, a positive correlation with total PFAS was also observed for octanoic and decanoic acids as well as for lysine and alanine. The authors note here that whilst no strong associations were found between MLCs and individual PFAS exposures, network analysis drew links between two PFASs and three of the MLCs. These associations were found to not be spurious in the network/correlation analysis (see Methods) in that they passed the non-rejection rate (NRR) threshold of 0.5 (see Methods). Firstly, perfluoropentanoic acid (PFPeA) associated with (1) MLC7 (certain ethyl phospholipids) with an inverse association of -0.11, NRR = 0.40 and also with (2) MLC9 (large (>57 carbon), polyunsaturated fatty acid (PUFA)-containing triglycerides) with an inverse association also of -0.11, NRR = 0.15). Secondly, PFTDA associated with (1) MLC7 (positive association: 0.1, NRR = 0.46) and also with (2) MLC8 (small (<50 carbon) short chain fatty acid-containing triacylglycerols, having a positive association: 0.14, NRR = 0.13 (associations visible in the network plot of **Figure 2**).

For a subset of the cohort (n = 116), detailed lifestyle data, including dietary data during pregnancy, were available, and this dataset was used to estimate dietary sources of PFAS (**Supplementary Figure 3**). Shellfish showed the strongest correlation with serum PFAS levels, with other food items also showing significant associations with the PFAS levels, such as fish, cereals and fruit juice. Fish oil consumption, which is not associated with PFAS exposure[34], increases serum levels of phospholipids and PUFA-containing TGs[35]. Conversely, PFAS exposure is associated with increased levels of these lipids[36]. This suggests that, to a large extent, seafood consumption drives maternal PFAS levels.

In agreement with previous studies[37, 38], we observed inverse correlations between the number of previous deliveries and levels of specific PFAS: PFOA (r = -0.44, FDR q < 0.01), PFHpS (r = -0.39, q < 0.01), PFOS (r = -0.31, q = 0.07), total PFAS (r = -0.26, q = 0.23, nominal p = 0.006).

Cord serum profiles in newborn infants associate with PFAS exposure

Partial correlation network analysis revealed a marked impact of maternal PFAS exposure on the cord serum metabolome of newborn infants (**Figure 2**). Inverse associations between cord serum lipids and PFOS, PFOA and total PFAS exposure were observed, particularly for clusters CLC2 (PUFA-containing phosphatidylcholines/PCs), CLC4 (lysophosphatidylcholines/LPCs) and CLC5 (sphingomyelin, abundant PCs). CMC4 (mainly specific amino acids) was positively associated with PFOS and perfluorodecanoic acid (PFDA) exposure.

Next, the infants were classified into four groups (quartiles) based on total maternal PFAS exposure levels during the pregnancy, as a sum of all measured PFAS exposures. Among the eight CLCs and four CMCs, one lipid cluster (CLC4) and one polar metabolite cluster (CMC4) were significantly different between the highest (Q4) and lowest (Q1) PFAS exposure quartiles, with two additional being of interest for their nominal p-value (CLC5 and CMC1) (**Supplementary Table 5**). At the individual metabolite level, several lipid species including LPCs, PCs, SMs and TGs were downregulated as total PFAS exposure increased. Specific amino acids, including phenylalanine, methionine and aspartic acid were significantly upregulated in the highest exposure group (**Figure 3**, **Supplementary Table 6**).

We then studied the associations between the matched maternal and cord serum metabolite levels using Spearman correlation. Following multiple hypothesis correction, only four lipids remained significantly correlated between the mothers and offspring. However, the correlation between the lipids was low overall ($|R| < 0.24$). Polar metabolites showed no significant correlations between maternal and cord blood samples.

Impact of PFAS exposure on cord serum lipids associated with risk of T1D progression

As the cord serum lipid profile associated with total maternal PFAS exposure here proved similar to that found previously as being associated with progression to T1D[8], we also examined the impact of PFAS exposure on T1D-associated lipids. First, we assigned the lipids from the present study to the same lipid clusters (LCs) as used in our previous study, and investigated their association with PFAS levels.

Of the ten lipid clusters used in the earlier study, five showed significant differences between the highest and lowest exposure groups (**Supplementary Table 7**). In our previous study, the most significantly-changing lipid clusters associated with T1D progression were LC2 (major PCs) and LC7 (sphingomyelins), which were down-regulated in newborn infants who later progressed to clinical T1D. In agreement with these results, the lipid levels in those same clusters were also reduced in the

highest (Q4) exposure group by comparison to the lowest (Q1) exposure levels in the current study. In addition, lipid clusters LC3 (LPCs) and LC6 (PUFA-containing phospholipids) showed clear differences in the current study, with lower lipid levels seen in the highest exposure group.

Of the 15 top-ranked lipids reported to have significant associations with T1D development[8], seven of these showed significant association with total PFAS exposure ($p < 0.05$). It is also of interest that the other lipids showed changes and also attained small nominal p-values ($p < 0.08$) (**Supplementary Table 8**). Among the individual PFAS, the strongest association was observed for perfluorononanoic acid (PFNA), with all 15 cord serum lipids were significantly associated with prenatal PFNA exposure.

Given the observed impact of prenatal exposure to PFAS on cord serum lipids associated with progression to T1D, we also examined whether HLA-conferred risk of T1D plays a role in mediating the impact of PFAS exposure on lipids in newborn infants. We divided the infants into two categories according to HLA-associated T1D risk: (low *vs.* increased; **Supplementary Table 1**) and two categories according to prenatal total PFAS exposure (quartiles 1 & 2 *vs.* 3 & 4). A multi-way ANOVA was then performed across these groups for the eleven lipids found associated with PFAS exposure. When examining the interaction effect between HLA risk and PFAS exposure, we found four lipids with a p-value < 0.05 (**Supplementary Table 9**).

In a subset of the cohort, which included 74 children enrolled in the clinical trial and for whom follow-up data were available, we then compared prenatal PFAS exposure in children who progressed to islet autoantibody positivity during the follow-up (n=10) *vs.* those that did not. We found that the children who progressed to one or more autoantibodies had elevated prenatal levels of several PFAS, with the largest effect being observed for PFOS, perfluoroheptane sulfonate (PFHpS) and perfluorohexane sulfonate (PFHxS) (**Supplementary Table 10**). We also compared the ratios of sphingomyelins and LPCs to PFAS between the two groups of children, as potential markers of the impact of exposure. Marked differences were observed between the groups, particularly for PFAS with 5 to 8 carbons in their structures (**Supplementary Table 10**).

Pre- and postnatal PFAS exposure in NOD mice alters offspring lipid profiles

Based on the metabolomics results from the mother-infant cohort, we hypothesized that PFAS exposure during pregnancy has a contributing, causal impact on phospholipid levels, which, in turn, associates with increased risk of T1D. Two previously-reported studies in NOD mice suggest that maternal PFAS exposure accelerates insulitis development and progression to autoimmune diabetes[29, 39]. We analyzed serum lipidomic profiles from these two studies, these being carried out in NOD mice (11 weeks of age) with exposure eitherto (1) PFUnDA in drinking water (0, 3, 30 and 300 μg/L)[29] or to (2) a mixture of persistent organic pollutants (POPs) in feeds (total PFAS intakes of 0, 0.14 (low), and 2.8μg/day (high)), with the low level corresponding to the approximate level of PFAS in human serum and the high level representing a 50-times higher level of total PFAS in serum[40, 41, 42]). It should be

noted that the levels of the other POPs were significantly lower than those of PFAS[41]. These exposures occurred at the times of mating, during gestation and lactation and until 11 weeks of age[39].

Exposure to PFUnDA caused significant changes in lipid profiles at the highest exposure level, with similar patterns of changes found with the two lower concentrations, although these did not reach statistical significance (**Supplementary Table 11**). Marked changes were observed also in the second study, where mice were exposed to the POP mixture, with the strongest effects seen in the high exposure group, but with significant changes still occurring also in the low exposure group which corresponded to expected human exposure levels (**Supplementary Table 12**). Specifically, exposure caused a marked reduction of a large number of phospholipids, with several PUFA-containing TGs being significantly down-regulated as well. We also identified significant changes in the levels of several free fatty acids, free cholesterol, amino acids, glycerol-3-phosphate and 3-hydroxybutyric acid, particularly in the high exposure group, with significant upregulation of TCA cycle metabolites.

As bile acids and PFAS utilize similar enterohepatic circulation[22, 23], we also examined the impact of the POP mixture on serum bile acid levels in serum of NOD mice. Indeed, the bile acid profiles were markedly altered, in a dose-dependent manner, due to the exposure to the POP mixture (**Figure 4a**; **Supplementary Table 13**). A majority of the bile acids, including the primary bile acids (CA, CDCA) were downregulated, while lithocholic acid (LCA) was markedly upregulated in comparison to the control group (**Figure 4a**; fold changes of 2.1 and 5.9 at low and high exposure to POP, p-values of 6.7×10^{-4} and 5.6×10^{-8}, respectively). Notably, there was a strong inverse association between the levels of LCA and the levels of SMs and LPCs (two examples shown in **Figure 4b-c**). Specifically, all SMs were downregulated (median R = -0.63, p = 0.000071-0.04) while 70% of the LPCs were significantly inversely associated with the LCA (median R = -0.60, p = 4.1×10^{-7}-0.01), except for LPC(22:3), which was positively correlated with LCA (R=0.54, p=0.007).

Next, we assigned the measured lipids to the same lipid clusters (LCs) as in our previous study[8] and investigated the association of PFAS exposure with these lipid clusters. Of the ten lipid clusters, four showed significant differences between the highest PFUnDA-exposure group and the control mice. One lipid cluster showed a significant difference even at the lowest level of exposure compared to control. In mice exposed to the POP mixture, eight of ten clusters showed significant changes between control and high exposure groups, and two clusters changed significantly between control and low exposure groups. In agreement with our previous study and our mother-infant cohort study presented here, LC2 decreased significantly with increasing PFAS exposure. In addition, clusters LC4, LC9 and LC10 showed clear differences in the current study with significantly lower lipid levels in the highest exposure group. These clusters contained mainly minor phospholipids, major TGs and long-chain PUFA-containing TGs. Using the cluster assignments from the earlier study[8], a remarkable similarity was observed when comparing the results across all four studies (**Figure 5**): (1) a previously-reported

study of the cord serum lipidome in relation to progression to T1D[8], (2) the association between PFAS exposure and cord serum lipid profiles in the mother-infant cohort presented here, and (3) the effects of PFUnDA and (4) a PFAS-containing POP mixture on the lipid profiles of NOD mice. Among the 15 individual lipids reported to have significant association with the development of T1D, 11 and 10 of these were detected also in the NOD mice, exposed to PFUnDA and POP mixture, respectively. Among these lipids, three showed significant differences between control and high exposure groups in the PFUnDA model and nine in the POP mixture exposure model. Notably, the pattern of the changes in the two mouse models were very similar, and *in vitro* exposure of macrophages to groups of chemicals from the POP mixture revealed that the PFAS mixture was driving the differences also in the POP mixture model[29, 39]. Strikingly, all lipid changes taking place in association with higher PFAS exposure occurred in the same direction as reported previously in relation to increased risk of T1D.

High exposure to PFOS in breastfed children at genetic risk for T1D associates with autoantibody positivity and with elevated levels of lithocholic acid

Next, we compared the plasma PFOS and lithocholic acid (LCA) concentration differences between the children who progressed to multiple islet autoantibodies (mAAb+) *vs.* controls (CTR), who remained AAb negative, in a previously-reported subgroup of the DIABIMMUNE study[11]. In line with our findings in the mother-child cohort, we observed higher level of PFOS, particularly in breast-fed (≥ 30 days) children who progressed to mAAb+ (n = 6) than in CTR (n = 20) at 18 months of age (p-value < 0.05, **Supplementary Figure 4a**). In addition, we found that in the longitudinal profile, PFOS remained persistently higher in the mAAb+ group than in CTR (**Supplementary Figure 4b**). We then sought to determine if LCA levels were altered with exposure. We found that children with the highest level (Q4) of PFOS exposure tended to have increased levels of LCA compared with children who had a low level (Q1) of PFOS exposure (**Supplementary Figure 5a**). We also observed that LCA differed between the cases and controls. Higher levels of LCA were found in children who progressed to mAAb+ than in the CTR at 6 and 36 months of age (p-value < 0.05, **Supplementary Figure 5b-c**).

DISCUSSION

By integrating PFAS exposure and metabolomic data from pregnant mothers with metabolomic data from their newborn infants, we were able to demonstrate altered cord serum metabolic signatures associated with high PFAS exposure during pregnancy and subsequently verify these in NOD mouse models of pre- and postnatal PFAS exposure. We also reported a remarkable similarity between the metabolic signature observed in the current (EDIA) study and the known signature associated with progression to T1D.

The composition of the cord blood metabolome reflects maternal metabolism, placental transfer across the maternal-fetal axis as well as fetal metabolism itself[43]. This may explain the weak associations between metabolic profiles of mothers and their offspring. The observed PFAS-associated

metabolic changes seen in cord blood were not associated with PFAS-related maternal metabolic changes. These fetal metabolic changes are therefore likely the result of PFAS exposure itself, rather than a downstream consequence of maternal metabolic changes. Several studies have shown that maternal levels of PFAS are reflected in the developing fetus[44] and there is a strong correlation between PFAS levels in maternal and cord blood[45]. One recent study indicates that PFAS concentrations in first trimester fetuses represent 5% to 27% of maternal plasma concentrations, fetal concentrations increasing with gestational age[46]. A comparison of transplacental transfer efficiency (TTE) for different PFAS suggests an inverse relationship with the chain-length of the perfluoroalkyl group and a somewhat lower transfer efficiency for perfluorosulfonic acids compared to perfluorocarboxylic acids[44]. We did not determine PFAS levels in newborn infants due to the limited volumes of samples available for quantification.

There is general consensus that exposure to PFOA and PFOS alters the immune system in experimental models, with documented effects including altered antibody and cytokine production[47]. In our study, we observed that prenatal PFAS exposure caused decreased levels of several phospholipids, particularly SMs and specific PCs, which were previously found to be persistently down-regulated in children who later progressed to islet autoimmunity[7] and clinical T1D[5, 6]. The importance of sphingolipid metabolism in the pathogenesis of T1D was recently highlighted by a genome-wide association study (GWAS) which identified eight gene polymorphisms involved in sphingolipid metabolism which contribute to T1D predisposition, and levels of which also correlated with the degree of islet autoimmunity in patients with recent-onset T1D[48]. Among the PFAS measured in our study, the main implicated drivers of the observed changes in cord serum phospholipid levels were PFNA, PFOS, PFUnDA and PFOA. Also, serine and palmitic acid (precursors of SMs) were found to be down-regulated with higher PFAS exposure and correlated with SM levels ($R > 0.4$), both in newborn infants as well as in NOD mice, where the exposure to PFAS was also associated with accelerated insulitis development[29]. We conclude that high PFAS exposure may alter sphingolipid levels during fetal development which may then go on to play a pathogenic role in the development of T1D later in life. The potential role of HLA-associated T1D risk in exacerbating the effect of prenatal PFAS exposure on lipid levels in the offspring, as suggested by our data, clearly demands further investigation.

Altered bile acid levels as observed in NOD mice exposed to POP mixtures, and in children positive for multiple islet autoantibodies, may explain altered lipid profiles. In animal models, LCA exposure has been shown to cause downregulation of LPCs and SMs in circulation[49], which is precisely what we have observed both in previous T1D studies (linking early lipid changes with progression to T1D later in life)[5, 6] as well as in the current study in relation to PFAS exposure. Notably, LCA has been previously linked to autoimmunity as it has been shown that LCA inhibits Th1 activation in Jurkat T cells, human and mouse primary CD4+ Th cells by inhibiting Th1 activation, mainly via vitamin D receptor[50]. LCA

is also an agonist for membrane receptor TGR5 which mediates the release of glucagon-like peptide 1 (GLP-1) promoting insulin release from pancreatic β-cells[21, 23]. Bile acid metabolism is also closely linked with gut microbiotal activity and, indeed, PFAS exposure has been shown to cause reduced microbiome diversity in infants[51]. In the DIABIMMUNE cohort, matching the sample studied here, the children that progressed to multiple islet autoantibodies later in life were found to have decreased alpha-diversity of gut microbiota. Decreased levels of LCA and increased levels of SMs in stool were also associated with relative overabundance of pathobionts, including mAAb+-associated *Ruminococcus*[11]. Our data presented in the current study thus suggests that PFAS impact absorption of bile acids, which may, in turn, affect circulating lipid levels (**Figure 6**).

We also observed an inverse association between maternal levels of specific PFAS and the number of previous deliveries, due to transfer of PFAS to the fetus, and excretion *via* breast milk (*i.e.*, breast feeding) which is in line with earlier reports[37, 38]. Interestingly, pooled analysis across multiple studies suggests that increasing birth order is associated with lower risk of T1D[52]. Our findings therefore support the notion that decreased maternal PFAS levels, due to multiple pregnancies, is one possible cause of this previously unexplained phenomenon.

Taken together, we conclude that high prenatal exposure to PFAS appears to alter lipid profiles in newborn infants, which, in turn, may increase the risk of islet autoimmunity and T1D. Our data also highlight a potential role for a gene-environment interaction (HLA risk genotype and prenatal PFAS exposure), which may lead to altered lipid profiles in newborn infants at-risk of developing T1D. Our findings may offer an explanation for the changing trend in the incidence of T1D in Western countries as well as underscore the need for investigations of how exposures to specific PFAS and other persistent chemical pollutants during pregnancy and early childhood affect the risk and pathogenesis of T1D.

METHODS

Mother-infant cohort

Pregnant women were recruited from January 28, 2013 to February 26, 2015, in the context of the EDIA (Early Dietary Intervention and Later Signs of Beta-Cell Autoimmunity: Potential Mechanisms) study, which is a small-scale intervention trial comparing weaning infants onto an extensively-hydrolyzed milk formula *vs*. a conventional cow's milk-based formula. Families were contacted at the time of the fetal ultrasonography visit, which is arranged for all pregnant women in Finland around gestational week 20. Written, informed consent was signed by the parents to permit analysis of their HLA genotype to exclude infants without HLA-conferred susceptibility to T1D. At this point, 68% of the infants to be born were excluded. Separate informed consent was obtained from eligible parents at the beginning of the third trimester to analyze the offspring's genotype and to continue in the intervention study.

The cord blood from 309 newborn infants was screened to determine the HLA genotype, as previously described[53]. The degree of HLA susceptibility to T1D was divided into six categories (high-risk, moderate-risk, low-risk, neutral, protective and strongly protective genotypes), as earlier defined[54]. A total of 89 infants were eligible for participation in the intervention study, carrying high-risk and moderate-risk genotypes. In that study, 82 infants were randomized and 73 remained in follow-up until the age of 12 months.

For the current study, the HLA risk categories were combined into two classes; the increased risk genotypes and the low-risk genotypes. Genotypes where HLA-(DR3)-DQA1*05-DQB*02 and/or DRB1*04:01/2/4/5-DQA1*03-DQB1*03:02 were present with each other, homozygous or heterozygous with a neutral haplotype were classified as increased risk and all other genotypes as low risk. Maternal diet during pregnancy was assessed by validated semiquantitative food frequency questionnaire[55]. Food and individual nutrient intakes were calculated using the national food composition database, Fineli[56]. We had access to 329 maternal serum samples collected at the beginning of the third trimester and 274 samples taken at delivery. We had, altogether, 300 cord blood samples. By pairing maternal and cord blood samples we obtained 264 paired mother-infant samples.

DIABIMMUNE study

The DIABIMMUNE study recruited 832 families in Finland (Espoo), Estonia (Tartu), and Russia (Petrozavodsk) with infants carrying HLA alleles that conferred risk for autoimmunity. The subjects involved in the current study were chosen from the subset (n = 62) of international DIABIMMUNE study children who progressed to multiple islet autoantibodies (n = 14) and controls (CTR, n = 38) who remained AAb negative in a longitudinal series of samples collected up to 3, 6, 12, 18, 24 and 36 months for each child[11]. The study groups were matched by HLA-associated diabetes risk, sex, country and period of birth. This study was conducted according to the guidelines in the Declaration of Helsinki. The Ethics and Research Committee of the participating Universities and Hospitals approved the study protocol. All families provided written informed consent prior to sample collection.

NOD mouse study – summary

The study setting of the two NOD mouse studies mice was reported previously[29, 39]. In short, NOD/ShiLtJ mice from the Jackson Laboratory (Maine, USA) were used for breeding at 8 and 10 weeks of age and randomly allocated to the exposure groups. Female offspring were, in both studies, exposed at mating, through gestation and early life until 11-12 weeks of age when the serum samples were collected, with 4-5 mice kept per cage and 5-8 mice per exposure group.

The exposure in the first study was to PFUnDA in the drinking water (n=8 per group) (0, 3, 30 and 300 µg/L, corresponding to 0.417, 4.17 and 41.7 µg/kg bw/day). The lowest exposure level of PFUnDA is about five times higher than the maximal calculated intake of PFOA in human infants. The exposure in the second study (n=5 per exposure group) was to a mixture of persistent organic pollutants in feed,

with a high and a low dose mixture (chemical composition based on human intake[39, 41]). The total intake of PFAS was 0.14 µg/day and 2.8 µg/day from the (1) low and (2) high dose groups, respectively, corresponding to (1) 1-50 times human serum levels of PFAS and (2) 20-1000 times the human serum levels[41]. In both studies, the mice had *ad libitum* access to food and water (Harlan Teklad 2919 irradiated, Madison, WI) and had a 12 h light/12 h dark cycle with 35–75% humidity.

All experiments were performed in conformity with the laws and regulations for experiments with live animals in Norway and were approved by the local representative of the Norwegian Animal Research Authority. In the NOD mouse model, insulitis is the most prominent feature preceding diabetes onset[57] with impaired macrophage phagocytosis being associated with seroconversion[58]. Insulitis was assessed by grading of hematoxylin and eosin-stained pancreatic tissue sections. Early signs of insulitis included an increased number of apoptotic cells, a decreased number of tissue-resident macrophages in pancreatic islets and reduced phagocytic function of macrophages isolated from the peritoneum.

Analysis of PFAS

Sample preparation and analysis for PFAS was carried out as described previously[59]. In short, 450 µL acetonitrile with 1% formic acid, and internal standards were added to 150 µL serum and samples subsequently treated with Ostro sample preparation in a 96-well plate for protein precipitation and phospholipid removal. The analysis of PFAS was performed using automated column-switching ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) (Waters, Milford, USA) using an ACQUITY C18 BEH 2.1×100mm×1.7µm column and a gradient with 30% methanol in 2mM NH4Ac water and 2mM NH4Ac in methanol with a flow rate of 0.3 mL/min. Quantitative analysis of the selected analytes was performed using the isotope dilution method; all standards (*i.e.*, internal standards, recovery standards, and native calibration standards) were purchased from Wellington Laboratories (Guelph, Ontario, Canada). The method's detection limits ranged between 0.02-0.19 ng/mL, depending on the analyte.

Analysis of bile acids

The bile acids were measured in NOD mice and in human serum as described recently[60], with some modifications in the sample preparation. 100 µL of acetonitrile and 10 µL of PFAS internal standard mixture (c = 200 µg/mL in methanol) and 20 µL of BA internal standard mixture (c = 440-670 µg/mL in methanol) and 50 µL NOD serum respectively were mixed, the samples were centrifuged and the organic phase was collected, evaporated to dryness after which ^{13}C injection standards were added (10 µL of 200 µg/mL PFAS in methanol) as was 300 µL of 2 mM NH4AC in water. For human samples, 20 µL of serum, using the same internal standard mixtures, was filtered through a frit filter plate (96-Well Protein Precipitation Filter Plate, Sigma Aldrich), and the effluent was collected and evaporated to dryness and the residue was dissolved in 20 µL of a 40:60 MeOH:H2O v/v mixture containing the same ^{13}C-PFAS injection standards. Analyses were performed on an ACQUITY UPLC system coupled to a

triple quadrupole mass spectrometer (Waters Corporation, Milford, USA) with an atmospheric electrospray interface operating in negative ion mode[60]. An external calibration with six calibration points (0.5-160 ng/mL), including a solvent blank, was carried out for use in quantitation.

Analysis of molecular lipids by UHPLC-QTOFMS

Serum samples were randomized and extracted using a modified version of the previously published Folch procedure[61]. In short, 10 µL of 0.9% NaCl and, 120 µL of CHCl3: MeOH (2:1, v/v) containing the internal standards (c = 2.5 µg/mL) was added to 10 µL of each serum sample. The standard solution contained the following compounds: 1,2-diheptadecanoyl-sn-glycero-3-phosphoethanolamine (PE(17:0/17:0)), N-heptadecanoyl-D-erythro-sphingosylphosphorylcholine (SM(d18:1/17:0)), N-heptadecanoyl-D-erythro-sphingosine (Cer(d18:1/17:0)), 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine (PC(17:0/17:0)), 1-heptadecanoyl-2-hydroxy-sn-glycero-3-phosphocholine (LPC(17:0)) and 1-palmitoyl-d31-2-oleoyl-sn-glycero-3-phosphocholine (PC(16:0/d31/18:1)), were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL, USA), and, triheptadecanoylglycerol (TG(17:0/17:0/17:0)) was purchased from Larodan AB (Solna, Sweden). The samples were vortex mixed and incubated on ice for 30 min after which they were centrifuged (9400 × g, 3 min). 60 µL from the lower layer of each sample was then transferred to a glass vial with an insert and 60 µL of CHCl3: MeOH (2:1, v/v) was added to each sample. The samples were stored at -80 °C until analysis.

Calibration curves using 1-hexadecyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine (PC(16:0e/18:1(9Z))), 1-(1Z-octadecenyl)-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine (PC(18:0p/18:1(9Z))), 1-stearoyl-2-hydroxy-sn-glycero-3-phosphocholine (LPC(18:0)), 1-oleoyl-2-hydroxy-sn-glycero-3-phosphocholine (LPC(18:1)), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (PE(16:0/18:1)), 1-(1Z-octadecenyl)-2-docosahexaenoyl-sn-glycero-3-phosphocholine (PC(18:0p/22:6)) and 1-stearoyl-2-linoleoyl-sn-glycerol (DG(18:0/18:2)), 1-(9Z-octadecenoyl)-sn-glycero-3-phosphoethanolamine (LPE(18:1)), N-(9Z-octadecenoyl)-sphinganine (Cer(d18:0/18:1(9Z))), 1-hexadecyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoethanolamine (PE(16:0/18:1)) from Avanti Polar Lipids, 1-Palmitoyl-2-Hydroxy-sn-Glycero-3-Phosphatidylcholine (LPC(16:0)), 1,2,3 trihexadecanoalglycerol (TG(16:0/16:0/16:0)), 1,2,3-trioctadecanoylglycerol (TG(18:0/18:0/18:)) and 3β-hydroxy-5-cholestene-3-stearate (ChoE(18:0)), 3β-Hydroxy-5-cholestene-3-linoleate (ChoE(18:2)) from Larodan, were prepared to the following concentration levels: 100, 500, 1000, 1500, 2000 and 2500 ng/mL (in CHCl3:MeOH, 2:1, v/v) including 1250 ng/mL of each internal standard.

The samples were analyzed by ultra-high-performance liquid chromatography quadrupole time-of-flight mass spectrometry (UHPLC-QTOFMS)[62]. Briefly, the UHPLC system used in this work was a 1290 Infinity II system from Agilent Technologies (Santa Clara, CA, USA). The system was equipped with a multi sampler (maintained at 10 °C), a quaternary solvent manager and a column thermostat

(maintained at 50 °C). Injection volume was 1 µL and the separations were performed on an ACQUITY UPLC® BEH C18 column (2.1 mm × 100 mm, particle size 1.7 µm) by Waters (Milford, MA, USA). The mass spectrometer coupled to the UHPLC was a 6545 QTOF from Agilent Technologies interfaced with a dual jet stream electrospray (Ddual ESI) ion source. All analyses were performed in positive ion mode and MassHunter B.06.01 (Agilent Technologies) was used for all data acquisition. Quality control was performed throughout the dataset by including blanks, pure standard samples, extracted standard samples and control serum samples. Relative standard deviations (% RSDs) for peak areas for lipid standards representing each lipid class in the control serum samples (n= 12) and in the pooled serum samples (n = 77) were calculated on average 15.9% and 13.6% (raw variation) in maternal samples and in cord blood samples, respectively. For serum samples from NOD mice, RSD was on average 11.9%. The lipid concentrations in pooled control samples showed % RSDs within accepted analytical limits at averages of 14.7% and 20.4% for the maternal and cord blood serum samples, respectively, and 7.3% for serum samples from NOD mice.

Mass spectrometry data processing was performed using the open source software package MZmine 2.18[63]. The following steps were applied in this processing: (1) Crop filtering with a m/z range of 350 – 1200 m/z and an RT range of 2.0 to 12 minutes, (2) Mass detection with a noise level of 750, (3) Chromatogram builder with a minimum time span of 0.08 min, minimum height of 1000 and a m/z tolerance of 0.006 m/z or 10.0 ppm, (4) Chromatogram deconvolution using the local minimum search algorithm with a 70% chromatographic threshold, 0.05 min minimum RT range, 5% minimum relative height, 1200 minimum absolute height, a minimum ration of peak top/edge of 1.2 and a peak duration range of 0.08 - 5.0, (5), Isotopic peak grouper with a m/z tolerance of 5.0 ppm, RT tolerance of 0.05 min, maximum charge of 2 and with the most intense isotope set as the representative isotope, (6) Peak filter with minimum 12 data points, a FWHM between 0.0 and 0.2, tailing factor between 0.45 and 2.22 and asymmetry factor between 0.40 and 2.50, (7) Join aligner with a m/z tolerance of 0.009 or 10.0 ppm and a weight for of 2, a RT tolerance of 0.1 min and a weight of 1 and with no requirement of charge state or ID and no comparison of isotope pattern, (8) Peak list row filter with a minimum of 10% of the samples (10) Gap filling using the same RT and m/z range gap filler algorithm with an m/z tolerance of 0.009 m/z or 11.0 ppm, (11) Identification of lipids using a custom database search with an m/z tolerance of 0.009 m/z or 10.0 ppm and a RT tolerance of 0.1 min, and (12) Normalization using internal standards PE(17:0/17:0), SM(d18:1/17:0), Cer(d18:1/17:0), LPC(17:0), TG(17:0/17:0/17:0) and PC(16:0/d30/18:1)) for identified lipids and closest ISTD for the unknown lipids followed by calculation of the concentrations based on lipid-class concentration curves.

An aliquot of each sample was collected and pooled and used as quality control sample, together with NIST SRM1950 reference plasma sample, an in-house pooled serum sample.

Analysis of polar metabolites by GC-TOFMS

Serum samples were randomized and sample preparation was carried out as described previously[64, 65]. In summary, 400 µL of MeOH containing ISTDs (heptadecanoic acid, deuterium-labeled DL-valine, deuterium-labeled succinic acid, and deuterium-labeled glutamic acid, c= 1 µg/mL) was added to 30 µl of the serum samples which were vortex mixed and incubated on ice for 30 min after which they were centrifuged (9400 × g, 3 min) and 350 µL of the supernatant was collected after centrifugation. The solvent was evaporated to dryness and 25 µL of MOX reagent was added and the sample was incubated for 60 min at 45 °C. 25 µL of MSTFA was added and after 60 min incubation at 45 °C 25 µL of the retention index standard mixture (n-alkanes, c=10 µg/mL) was added.

The analyses were carried out on an Agilent 7890B GC coupled to 7200 Q-TOF MS. Injection volume was 1 µL with 100:1 cold solvent split on PTV at 70 °C, heating to 300 °C at 120 °C/minute. Column: Zebron ZB-SemiVolatiles. Length: 20m, I.D. 0.18mm, film thickness: 0.18 µm. With initial Helium flow 1.2 mL/min, increasing to 2.4 mL/min after 16 mins. Oven temperature program: 50 °C (5 min), then to 270°C at 20 °C/min and then to 300 °C at 40 °C/min (5 min). EI source: 250 °C, 70 eV electron energy, 35µA emission, solvent delay 3 min. Mass range 55 to 650 amu, acquisition rate 5 spectra/s, acquisition time 200 ms/spectrum. Quad at 150 °C, 1.5 mL/min N_2 collision flow, aux-2 temperature: 280 °C.

Calibration curves were constructed using alanine, citric acid, fumaric acid, glutamic acid, glycine, lactic acid, malic acid, 2-hydroxybutyric acid, 3-hydroxybutyric acid, linoleic acid, oleic acid, palmitic acid, stearic acid, cholesterol, fructose, glutamine, indole-3-propionic acid, isoleucine, leucine, proline, succinic acid, valine, asparagine, aspartic acid, arachidonic acid, glycerol-3-phosphate, lysine, methionine, ornithine, phenylalanine, serine and threonine purchased from Sigma-Aldrich (St. Louis, MO, USA) at concentration range of 0.1 to 80 µg/mL. An aliquot of each sample was collected and pooled and used as quality control samples, together with a NIST SRM 1950 serum sample and an in-house pooled serum sample. Relative standard deviations (% RSDs) of the metabolite concentrations in control serum samples showed % RSDs within accepted analytical limits at averages of 12.3% and 19.6% for the maternal and cord blood serum samples, respectively, and 7.2% for serum samples from NOD mice.

Analysis of islet autoantibodies (EDIA and DIABIMMUNE studies)

Four diabetes-associated autoantibodies were analyzed from each serum sample with specific radiobinding assays: insulin autoantibodies (IAA), glutamic acid decarboxylase antibodies (GADA), islet antigen-2 antibodies (IA-2A), and zinc transporter 8 antibodies (ZnT8A) as described previously[66]. Islet cell antibodies (ICA) were analyzed with immunofluoresence in those subjects who tested positive for at least one of the biochemical autoantibodies. The cut-off values were based on the 99th percentile in non-diabetic children and were 2.80 relative units (RU) for IAA, 5.36 RU for GADA, 0.78 RU for IA-2A and 0.61 RU for ZnT8A, Detection limit in the ICA assay was 2.5 Juvenile Diabetes Foundation units (JDFU).

Data transformation and descriptive statistical analysis

Spreadsheets containing concentration data were converted to .csv format for loading into the R statistical programming language[67]. For all datasets, the following transformations were carried out:

1. NA values in the data were replaced with zeroes.
2. The aforementioned zeroes were then replaced with imputed half-minimums (for each variable, the minimum value was found, and half of this value was used).
3. All values were log2 transformed.
4. Each variable was scaled to zero mean and unit variance.

Total PFAS exposure in pregnant mothers was assessed as a simple sum of exposure to all measured PFAS. Mothers and matching children were then sorted into quartiles of this total maternal PFAS exposure. Analysis by both ANOVA and Tukey's honest significant difference (TukeyHSD) were carried out on all lipids and metabolites in the infants' cord blood, grouping these into the aforementioned maternal PFAS exposure quartiles to reveal any significant changes in their level as exposure increased. To visualize this, the R beanplot package (version 1.2)[68] was used to show both the changes in the means across exposure quartiles, and the densities of samples in each quartile.

Multiu-way analysis of variance was performed with factors HLA risk and PFAS exposure) and their interactions in MATLAB R2017b (Mathworks, Inc., Natick, MA, USA) using Statistical Toolbox.

The Wilcoxon rank-sum test was used in comparing the two study groups of samples (*e.g.* CTR *vs.* mAAb+ group) in a specific age cohort. These statistical analyses were computed in MATLAB 2017b using the statistical toolbox. For statistical comparison, subjects with missing peaks for the given quantified compound and children who were not exclusively breast-fed for 30 days were excluded. The longitudinal profiles of the PFOS in the samples obtained from children who progressed to multiple autoantibodies and autoantibody negative controls were compared using linear mixed-effects model with the fixed effect being case, age, sex, and the random effect being subject-wise variation using lme4 package in R. The fully parametrized model was compared with a null model using analysis of variance (ANOVA) as deployed in the lme4 package. The locally weighted regression plot was made using smoothing interpolation function loess (with span = 0.85) available from ggplot2 package in R. The individual metabolite levels were visualized as scatter plot as well as box plot using GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA).

Clustering

All metabolomics datasets were then analyzed using the mclust R package (version 5.4.1)[32] to assign variables (lipids / metabolites) from each dataset to separate clusters. Here, mclust attempts to fit various model types and assesses their performance using the Bayesian Information Criterion (BIC).

Maximization of the BIC is a gold-standard method for model selection, particularly useful in the case of clustering of data, to choose the optimum number of clusters by way of maximizing the likelihood of fit of the clustering model, whilst this metric penalizes (and thereby avoids) unnecessary complexity. As the mclust package tries its various model types and increasing numbers of clusters into which to divide the data. The BIC is calculated at each iteration, and the optimal (maximal) BIC will occur when the lowest number of clusters are used before the point at which increasing the number of clusters gives a lesser return in the fitness of the model. The highest BIC achieved by mclust for each dataset was therefore used to determine both the model type and the number of clusters into which the variables should be divided. The variables in each dataset were accordingly given numbers to denote their cluster membership. Plots demonstrating the distribution of BICs for each dataset, and therefore justifying the choice of a specific number of clusters for downstream processing, are given in supplementary data (**Supplementary Figure 1).**

For each sample in each dataset, cluster variables were then generated. For each dataset, the number of clusters *k* found by mclust equals the number of cluster variables generated. Each cluster variable is the mean value of the lipids / metabolites that make up that cluster, meaning that samples in that dataset are represented now only by *k* values. These cluster variables were given acronyms indicating the dataset from which they were generated (Maternal blood Lipids Cluster – MLC, Maternal polar Metabolites Cluster - MMC, Offspring Lipids Cluster - CLC, Offspring polar Metabolites Cluster – CMC).

These cluster variable acronyms had their cluster numbers (1-*k* for each dataset) appended to them to form their final labels. Assignment of individual lipids / metabolites to each cluster variable is given for each dataset in supplementary files, along with total membership counts in each cluster. PFAS compounds were not clustered in this manner and so retain their names.

Identification of potentially-confounded variables in the mother-child cohort

Analyses which draw upon large numbers of human samples in order to elucidate patterns therein can suffer from the effects of confounding between their features of biological / experimental interest (in this case, the levels of circulating lipids and polar metabolites) and the underlying distribution of the demographics of that cohort. It is possible, in the case of our analysis, for example, that the results that we see in the form of lipid/metabolites levels and therefore the downstream associations that we find between lipids / metabolites and clinical outcomes are, in reality, attributable to a greater or lesser degree to anthropometric values such as the age, sex or BMI of the subjects in the cohort. In the event of significant confounding effects of these demographic variables upon a sizeable number of lipids / polar metabolites, various correction methods would need to be employed to mitigate such effects, if possible.

In order to rule out the effects of these common, but important, demographic factors, a comprehensive analysis was carried out, modeling the relationship between all lipids and polar metabolites in both mothers and children, and the variables of age, BMI and sex. Relationships between lipids / polar metabolites and the covariates of age / BMI / sex were modeled applying generalized linear models (GLMs) using the glm() package provided as a base function in the R statistical programming language. Models were constructed aiming to link each lipid/polar metabolite with the covariates of age, BMI and sex, returning coefficients, intercepts and, crucially significance values for any relationship that could be modeled. For each dataset (n = 4 (maternal lipids, maternal polar metabolites, offspring lipids, offspring polar metabolites)), all p-values of associations between the covariates and individual lipids/polar metabolites were collected. These were then corrected for multiple hypothesis generation, both by Bonferroni and Benjamini-Hochberg correction methods.

This analysis therefore attempted to fit a total of (n = (206 maternal lipids + 206 offspring lipids + 34 maternal polar metabolites + 31 offspring metabolites) * 3 covariates = **1431**) models by GLM, thereby exhaustively searching for any confounded relationships. The largest number (n = 46) of significant relationships was found, intuitively, to occur between maternal blood lipids and BMI, although none of these potential associations remained after multiple hypothesis correction. In fact, the only associations which did pass multiple hypothesis correction were two (with Bonferroni correction) or three (with Benjamini-Hochberg correction) associations between BMI and maternal polar metabolites, accounting for only between 5.9% to 8.8% of the maternal polar metabolites. No other datasets had any significant relationship between their lipids/polar metabolites and the potentially-confounding factors of age, BMI or sex. In summary, therefore, a maximum of three relationships from a potential pool of 1431 could be argued to be affected by age, BMI or sex in our data, corresponding to 0.21%. Therefore (1) the data itself does not suffer from confounding by the covariates of age, BMI or sex, before any downstream analysis was performed, thus removing this as a concern for all downstream analyses and inferences.

A summary of the numbers of significant relationships in all datasets and regarding all three covariates (age, BMI, sex) are given both as p and corrected q values in **Supplementary Table 14**.

Partial correlation network analysis

With dataset dimensionality reduced to the aforementioned cluster variables, partial correlation analysis was employed, taking all cluster variables into account, along with clinical variables, simultaneously. Pairwise Spearman correlations between all of the aforementioned variables were calculated. To subsequently visualize this, the corrplot R package (version 0.84) was used. For legibility of figures, the colors of these plots generated by corrplot were limited to either solid orange or blue for positive or inverse correlations respectively, with correlation strength represented purely through the size of the filled circles for each pairwise correlation.

For network analysis and representation based on the generated partial correlations, the qpNrr() function from the qpgraph R package (version 2.16.0)[33] was run with default parameters to estimate non-rejection rates (NRRs) of the aforementioned correlation matrix of all datasets' cluster variables and clinical features. This is a means for rejecting spurious correlations. The obtained NRR matrix was then filtered at various thresholds (0.1 to 1, with an increment of 0.1) to provide the edges for network graphing. The distribution of NRRs was also visualized as a histogram to assist with choosing an appropriate cut-off threshold for retaining plausible, and rejecting likely spurious, correlations. These are given in supplementary figures. Based on the distribution of NRRs and observed network topology of the generated networks, a cut-off of 0.5 was deemed appropriate.

The Rgraphviz R package (version 2.26.0)[69] was used to generate network topography plots. Node and edge properties for these network graphs were generated in a custom fashion. Edges were colored by the directionality of the relationship between the nodes that they connect. Edge width was plotted as a function of the strength of the Spearman correlation between the two variables that the edge connects. Nodes were colored and shaped purely for clarity and to group like variables (PFAS, clinical variables, cluster variables) and sample sources (mothers, infants) together. Network layout is generated by the Rgraphviz package itself, and layout was set to the "neato" parameter to balance clarity and compactness.

For both the final correlation plot and network figure, values were used only from those sample identifiers uniquely represented in all four datasets (maternal blood lipids, maternal polar metabolites, offspring blood lipids, offspring blood polar metabolites), totaling (n = 224 samples, n = 59 variables).

Data accessibility

The lipidomics datasets and the clinical metadata generated in this study were submitted to MetaboLights[70] under accession number (MTBLS875). The relevant clinical metadata from the EDIA study was linked to the lipidomics dataset using the ISA-creator package from MetaboLights.

Ethical approval and informed consent

The Ethical Committee of the Pirkanmaa Hospital District has approved the EDIA study. The parents gave their written informed consent to their own participation and to the participation of their newborn infant in the study. The conduct of the EDIA study conforms to the standards of the Declaration of Helsinki.

Acknowledgments

The EDIA study was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (No. 1DP3DK094338-01 to M.K.), the Academy of Finland Centre of Excellence in Molecular Systems Immunology and Physiology Research 2012-17, No. 250114 to M.K. and M.O.), Academy of Finland postdoctoral grant (No. 323171 to S.L.) and the Medical

Research Funds, Tampere and Helsinki University Hospitals (to M.K.). The current work was also supported by the Juvenile Diabetes Research Foundation (2-SRA-2016-341-S-B to M.K., T.H. and M.O., 2-SRA-2016-289-S-B to M.O.), Academy of Finland (No. 292568 to M.O.), Vetenskapsrådet (No. 2016-05176 to T.H.), and the Knowledge Foundation (to T.H.). The authors thank Samira Salihović and Tuomas Lindeman for technical expertise in the development of bile acid analysis method. Bile acid analyses were partly performed at the Turku Metabolomics Centre, core facility that is part of the Biocenter Finland national research infrastructure.

Author contributions

T.H. initially proposed the study of PFAS exposure in the mother-infant study (EDIA), while M.O. proposed the follow-up metabolomics studies in NOD mice. T.H., M.O. and M.K. initiated and designed the study. M.K. is the principal investigator of the mother-infant study. M.O. supervised data analysis and integration, and together with T.H. did the primary interpretation of analytical outcomes. A.M., M.O., S.L. and T.H. analyzed the data. T.H. supervised metabolomic and PFAS analyses. T.S., D.G., C.C., D.D. and A.D. performed sample analyses. J.B., H.D. and U.C.N. conducted the studies in NOD mice, while H.F.B. and K.Z. assembled the internal relationship of chemicals in the feed for POP-exposure. H.S. contributed clinical research in the mother-infant cohort. J.I. performed genetic analyses in the mother-infant study. S.M.V. performed the study of dietary intake in the mother-infant study. A.M., M.O. and T.H. wrote the first version of the paper. All authors approved the final version.

REFERENCES

1. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet* **383**, 69-82 (2014).

2. Noble JA, Erlich HA. Genetics of type 1 diabetes. *Cold Spring Harb Perspect Med* **2**, a007732 (2012).

3. Achenbach P, Bonifacio E, Koczwara K, Ziegler AG. Natural history of type 1 diabetes. *Diabetes* **54 Suppl 2**, S25-31 (2005).

4. Knip M, Veijola R, Virtanen SM, Hyöty H, Vaarala O, Åkerblom HK. Environmental triggers and determinants of type 1 diabetes. *Diabetes* **54**, S125-136 (2005).

5. Orešič M, *et al.* Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. *J Exp Med* **205**, 2975-2984 (2008).

6. Lamichhane S, *et al.* Dynamics of Plasma Lipidome in Progression to Islet Autoimmunity and Type 1 Diabetes - Type 1 Diabetes Prediction and Prevention Study (DIPP). *Sci Rep* **8**, 10635 (2018).

7. Johnson RK, *et al.* Metabolite-related dietary patterns and the development of islet autoimmunity. *Sci Rep* **9**, 14819 (2019).

8. Oresic M, *et al.* Cord serum lipidome in prediction of islet autoimmunity and type 1 diabetes. *Diabetes* **62**, 3268-3274 (2013).

9. La Torre D, *et al.* Decreased cord-blood phospholipids in young age-at-onset type 1 diabetes. *Diabetes* **62**, 3951-3956 (2013).

10. Velagapudi VR, *et al.* The gut microbiota modulates host energy and lipid metabolism in mice. *J Lipid Res* **51**, 1101-1112 (2010).

11. Kostic AD, *et al.* The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* **17**, 260-273 (2015).

12. Vatanen T, *et al.* The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature* **562**, 589-594 (2018).

13. Patterson CC, Dahlquist GG, Gyürüs E, Green A, Soltész G. Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet* **373**, 2027-2033 (2009).

14. Berhan Y, Waernbaum I, Lind T, Möllsten A, Dahlquist G, for the Swedish Childhood Diabetes Study G. Thirty Years of Prospective Nationwide Incidence of Childhood Type 1 Diabetes: The Accelerating Increase by Time Tends to Level Off in Sweden. *Diabetes* **60**, 577-581 (2011).

15. Harjutsalo V, Sund R, Knip M, Groop PH. Incidence of type 1 diabetes in Finland. *JAMA* **310**, 427-428 (2013).

16. Kaikkonen R, *et al.* Health and well-being inequalities among children and their families. (ed^(eds). National Institute for Health and Wellfare (2012).

17. Wang T, Vestergren R, Herzke D, Yu J, Cousins IT. Levels, Isomer Profiles, and Estimated Riverine Mass Discharges of Perfluoroalkyl Acids and Fluorinated Alternatives at the Mouths of Chinese Rivers. *Environ Sci Technol* **50**, 11584-11592 (2016).

18. Gebbink WA, Glynn A, Berger U. Temporal changes (1997-2012) of perfluoroalkyl acids and selected precursors (including isomers) in Swedish human serum. *Environ Pollut* **199**, 166-173 (2015).

19. Bjerregaard-Olesen C, *et al.* Time trends of perfluorinated alkyl acids in serum from Danish pregnant women 2008-2013. *Environ Int* **91**, 14-21 (2016).

20. Croes K, *et al.* Persistent organic pollutants (POPs) in human milk: A biomonitoring study in rural areas of Flanders (Belgium). *Chemosphere* **89**, 988-994 (2012).

21. Prawitt J, Caron S, Staels B. Bile acid metabolism and the pathogenesis of type 2 diabetes. *Curr Diab Rep* **11**, 160-166 (2011).

22. Zhao W, *et al.* Na+/Taurocholate Cotransporting Polypeptide and Apical Sodium-Dependent Bile Acid Transporter Are Involved in the Disposition of Perfluoroalkyl Sulfonates in Humans and Rats. *Toxicol Sci* **146**, 363-373 (2015).

23. Chiang JY. Recent advances in understanding bile acid homeostasis. *F1000Res* **6**, 2029 (2017).

24. Fujii Y, *et al.* Toxicokinetics of perfluoroalkyl carboxylic acids with different carbon chain lengths in mice and humans. *J Occup Health* **57**, 1-12 (2015).

25. Harada KH, *et al.* Biliary excretion and cerebrospinal fluid partition of perfluorooctanoate and perfluorooctane sulfonate in humans. *Environ Toxicol Pharmacol* **24**, 134-139 (2007).

26. Matilla-Santander N, *et al.* Exposure to Perfluoroalkyl Substances and Metabolic Outcomes in Pregnant Women: Evidence from the Spanish INMA Birth Cohorts. *Environ Health Perspect* **125**, 117004 (2017).

27. Bodin J, Stene LC, Nygaard UC. Can exposure to environmental chemicals increase the risk of diabetes type 1 development? *Biomed Res Int* **2015**, 208947 (2015).

28. Liu S, Yin N, Faiola F. PFOA and PFOS Disrupt the Generation of Human Pancreatic Progenitor Cells. *Environmental Science & Technology Letters* **5**, 237-242 (2018).

29. Bodin J, Groeng EC, Andreassen M, Dirven H, Nygaard UC. Exposure to perfluoroundecanoic acid (PFUnDA) accelerates insulitis development in a mouse model of type 1 diabetes. *Toxicol Rep* **3**, 664-672 (2016).

30. Predieri B, Iughetti L, Guerranti C, Bruzzi P, Perra G, Focardi SE. High Levels of Perfluorooctane Sulfonate in Children at the Onset of Diabetes. *Int J Endocrinol* **2015**, 234358 (2015).

31. Land M, *et al.* What is the effect of phasing out long-chain per- and polyfluoroalkyl substances on the concentrations of perfluoroalkyl acids and their precursors in the environment? A systematic review. *Environmental Evidence* **7**, 4 (2018).

32. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J* **8**, 289-317 (2016).

33. Castelo R, Roverato A. Reverse engineering molecular regulatory networks from microarray data with qp-graphs. *J Comp Biol* **16**, 213-227 (2009).

34. Averina M, Brox J, Huber S, Furberg AS. Perfluoroalkyl substances in adolescents in northern Norway: Lifestyle and dietary predictors. The Tromso study, Fit Futures 1. *Environ Int* **114**, 123-130 (2018).

35. Ottestad I, *et al.* Fish oil supplementation alters the plasma lipidomic profile and increases long-chain PUFAs of phospholipids and triglycerides in healthy subjects. *PLoS One* **7**, e42550 (2012).

36. Salihovic S, *et al.* Identification of metabolic profiles associated with human exposure to perfluoroalkyl substances. *J Expo Sci Environ Epidemiol* **29**, 196-205 (2019).

37. Tsai MS, *et al.* Determinants and Temporal Trends of Perfluoroalkyl Substances in Pregnant Women: The Hokkaido Study on Environment and Children's Health. *Int J Environ Res Public Health* **15**, (2018).

38. Lauritzen HB, *et al.* Factors Associated with Maternal Serum Levels of Perfluoroalkyl Substances and Organochlorines: A Descriptive Study of Parous Women in Norway and Sweden. *PLoS One* **11**, e0166127 (2016).

39. Berntsen HF, *et al.* Decreased macrophage phagocytic function due to xenobiotic exposures in vitro, difference in sensitivity between various macrophage models. *Food Chem Toxicol* **112**, 86-96 (2018).

40. Gebbink WA, Glynn A, Darnerud PO, Berger U. Perfluoroalkyl acids and their precursors in Swedish food: The relative importance of direct and indirect dietary exposure. *Environ Pollut* **198**, 108-115 (2015).

41. Berntsen HF, Berg V, Thomsen C, Ropstad E, Zimmer KE. The design of an environmentally relevant mixture of persistent organic pollutants for use in in vivo and in vitro studies. *J Toxicol Environ Health A* **80**, 1002-1016 (2017).

42. Haug LS, Thomsen C, Becher G. Time trends and the influence of age and gender on serum concentrations of perfluorinated compounds in archived human samples. *Environ Sci Technol* **43**, 2131-2136 (2009).

43. Hart K, *et al.* Time trends and transplacental transfer of perfluorinated compounds in melon-headed whales stranded along the Japanese coast in 1982, 2001/2002, and 2006. *Environ Sci Technol* **42**, 7132-7137 (2008).

44. Winkens K, Vestergren R, Berger U, Cousins IT. Early life exposure to per- and polyfluoroalkyl substances (PFASs): A critical review. *Emerging Contaminants* **3**, 55-68 (2017).

45. Kim S, *et al.* Trans-placental transfer of thirteen perfluorinated compounds and relations with fetal thyroid hormones. *Environ Sci Technol* **45**, 7465-7472 (2011).

46. Mamsen LS, *et al.* Concentration of perfluorinated compounds and cotinine in human foetal organs, placenta, and maternal plasma. *Sci Total Environ* **596-597**, 97-105 (2017).

47. DeWitt JC, *et al.* Immunotoxicity of perfluorooctanoic acid and perfluorooctane sulfonate and the role of peroxisome proliferator-activated receptor alpha. *Crit Rev Toxicol* **39**, 76-94 (2009).

48. Holm LJ, *et al.* Abnormal islet sphingolipid metabolism in type 1 diabetes. *Diabetologia* **61**, 1650-1661 (2018).

49. Matsubara T, Tanaka N, Patterson AD, Cho JY, Krausz KW, Gonzalez FJ. Lithocholic acid disrupts phospholipid and sphingolipid homeostasis leading to cholestasis in mice. *Hepatology* **53**, 1282-1293 (2011).

50. Pols TWH, Puchner T, Korkmaz HI, Vos M, Soeters MR, de Vries CJM. Lithocholic acid controls adaptive immune responses by inhibition of Th1 activation through the Vitamin D receptor. *PLoS One* **12**, e0176715 (2017).

51. Iszatt N, *et al.* Environmental toxicants in breast milk of Norwegian mothers and gut bacteria composition and metabolites in their infants at 1 month. *Microbiome* **7**, 34 (2019).

52. Cardwell CR, *et al.* Birth order and childhood type 1 diabetes risk: a pooled analysis of 31 observational studies. *Int J Epidemiol* **40**, 363-374 (2011).

53. Hermann R, *et al.* HLA DR-DQ-encoded genetic determinants of childhood-onset type 1 diabetes in Finland: an analysis of 622 nuclear families. *Tissue Antigens* **62**, 162-169 (2003).

54. Ilonen J, *et al.* Genetic susceptibility to type 1 diabetes in childhood - estimation of HLA class II associated disease risk and class II effect in various phases of islet autoimmunity. *Pediatr Diabetes* **17 Suppl 22**, 8-16 (2016).

55. Erkkola M, Karppinen M, Javanainen J, Rasanen L, Knip M, Virtanen SM. Validity and reproducibility of a food frequency questionnaire for pregnant Finnish women. *Am J Epidemiol* **154**, 466-476 (2001).

56. National Institute of Health and Welfare. National Food Composition Database in Finland (https://fineli.fi/fineli/en/index). (ed^(eds).

57. Mullen Y. Development of the Nonobese Diabetic Mouse and Contribution of Animal Models for Understanding Type 1 Diabetes. *Pancreas* **46**, 455-466 (2017).

58. O'Brien BA, *et al.* A deficiency in the in vivo clearance of apoptotic cells is a feature of the NOD mouse. *J Autoimmun* **26**, 104-115 (2006).

59. Salihovic S, Karrman A, Lindstrom G, Lind PM, Lind L, van Bavel B. A rapid method for the determination of perfluoroalkyl substances including structural isomers of perfluorooctane sulfonic acid in human serum using 96-well plates and column-switching ultra-high performance liquid chromatography tandem mass spectrometry. *J Chromatogr A* **1305**, 164-170 (2013).

60. Salihovic S, *et al.* Simultaneous Determination of Per- and Polyfluoroalkyl Substances and Bile Acids in Human Serum Using Ultra-High-Performance Liquid Chromatography-Tandem Mass Spectrometry. *chemRxiv*, 9772499 (2019).

61. Pedersen HK, *et al.* A computational framework to integrate high-throughput '-omics' datasets for the identification of potential mechanistic links. *Nat Protoc* **13**, 2781-2800 (2018).

62. Nygren H, Seppanen-Laakso T, Castillo S, Hyotylainen T, Oresic M. Liquid chromatography-mass spectrometry (LC-MS)-based lipidomics for studies of body fluids and tissues. *Methods Mol Biol* **708**, 247-257 (2011).

63. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).

64. Hartonen M, Mattila I, Ruskeepää A-L, Orešič M, Hyötyläinen T. Characterization of cerebrospinal fluid by comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry. *J Chromatogr A* **1293**, 142-149 (2013).

65. Castillo S, Mattila I, Miettinen J, Orešič M, Hyotylainen T. Data analysis tool for comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. *Anal Chem* **83**, 3058-3067 (2011).

66. Knip M, *et al.* Prediction of type 1 diabetes in the general population. *Diabetes Care* **33**, 1206-1212 (2010).

67. R Development Core Team. R: A language and environment for statistical computing. (ed^(eds). R Foundation for Statistical Computing (2018).

68. Kampstra P. Beanplot: a boxplot alternative for visual comparison of distributions. *J Stat Soft* **28**, 1-9 (2008).

69. Hansen KD, *et al.* Rgraphviz: Provides plotting capabilities for R graph objects. R package version 2.27.0. (2018).

70. Kale NS, *et al.* MetaboLights: An Open-Access Database Repository for Metabolomics Data. *Curr Protoc Bioinformatics* **53**, 14 13 11-18 (2016).

Figure captions

Figure 1. Overview of the workflow integrating prenatal PFAS exposure assessment, serum metabolomics and risk of type 1 diabetes. **a**. In a mother-infant cohort, PFAS levels and metabolomic profiles were determined from pregnant mothers, and metabolomics performed on cord serum from newborn infants. **b**. Metabolites were summarized as clusters, and (**c**) associations between prenatal PFAS exposure and metabolomes were studied. Cord serum lipid changes due to prenatal PFAS exposure were then compared to (**d**) previously reported[8] lipid-related differences between newborn infants who progressed to T1D later in life *vs.* those that remained healthy (Type 1 Diabetes Prediction and Prevention study – DIPP), and to (**e**) changes in lipid profiles brought on by exposure to a single PFAS compound or mixture of persistent organic pollutants, respectively, from two studies in non-obese diabetic (NOD) mice. **f**. The data across the four different studies (**a**, **d**, **e**) were summarized and compared by assigning lipids from each respective study to lipid clusters from the T1D study[8].

Figure 2. Partial correlation network showing associations between demographic data, maternal PFAS levels and lipidome / metabolome cluster variables from mothers and their newborn infants. The network was constructed using the qpgraph R package[33]. Node color represents different types of variables, edge color denotes a positive (orange) or negative (blue) association. The threshold non-rejection rate was set as 0.5. Node abbreviations: PregWk, weeks of pregnancy; HLA, HLA risk locus (1 = lower risk, 2 = higher risk); Head_C, head circumference of child; BWei, birth weight; BLen, birth length; MBMI, maternal BMI; MAge, maternal age.

Figure 3. Beanplots showing levels of four selected lipids and two polar metabolites, measured in cord serum, having significant (adjusted $p < 0.05$) associations with total maternal prenatal PFAS exposure (**Supplementary Table 6**). X-axis numbers correspond to total maternal PFAS exposure level from lowest (1) to highest (4) quartile. Red, horizontal bars indicate means, black horizontal bars are individual sample values and "bean" width represents the density of samples occurring at that level. All values plotted are log2 transformed and scaled to zero mean and unit variance.

Figure 4. Impact of POP mixture on levels of lithocholic acid in NOD mice. **a.** LCA levels in NOD mice exposed to POP mixture at two different doses. Associations of LCA levels in NOD mice exposed to POP mixture, with **b.** SM(d38:2) and **c.** LPC(22:5).

Figure 5. Comparison of lipidomic profiles across four different studies, using lipid cluster assignments from an earlier study[8]. **a**. Cord serum profiles from progressors to T1D (yellow bars) and control children (blue bars), from a previous report in the Diabetes Prediction and Prevention (DIPP) study in Finland[8]. **b**. Cord serum from mother-infant (EDIA) cohort, with high PFAS exposure (yellow) and low exposure (blue). **c**. NOD mice exposed to a high level of PFUnDA (yellow) and unexposed control mice (blue). **d**. NOD mice exposed to POP mixture at a high level (yellow) and unexposed mice

(blue). **e**. Fold changes between the groups in **a**-**d** of lipids in cluster LC2. Statistical significance levels: *p<0.05, **p<0.01, ***p<0.001.

Figure 6. Summary of findings, linking prenatal PFAS exposure with decreased phospholipids and increased risk of T1D in the offspring. PFAS exposure decreases levels of most bile acids, which may be due to common enterohepatic circulation, but the microbially-produced secondary bile acid LCA is markedly increased. This proinflammatory bile acid is known to decrease SMs and LPCs in circulation, which is also supported by our data. Phospholipids (incl. SMs) are known to be inversely associated with risk of T1D[5, 6, 7, 8, 9].

FIGURE 1

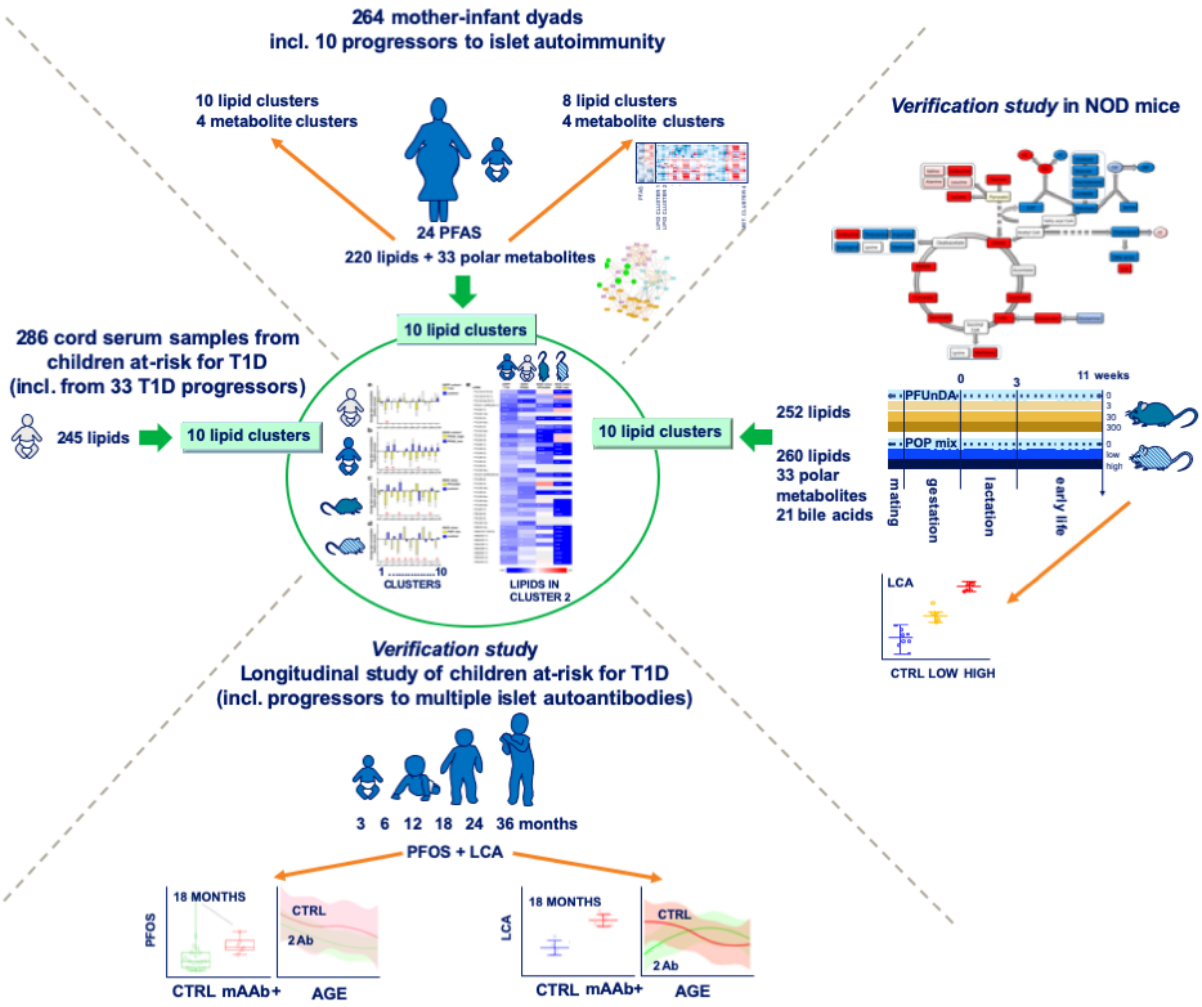

FIGURE 2

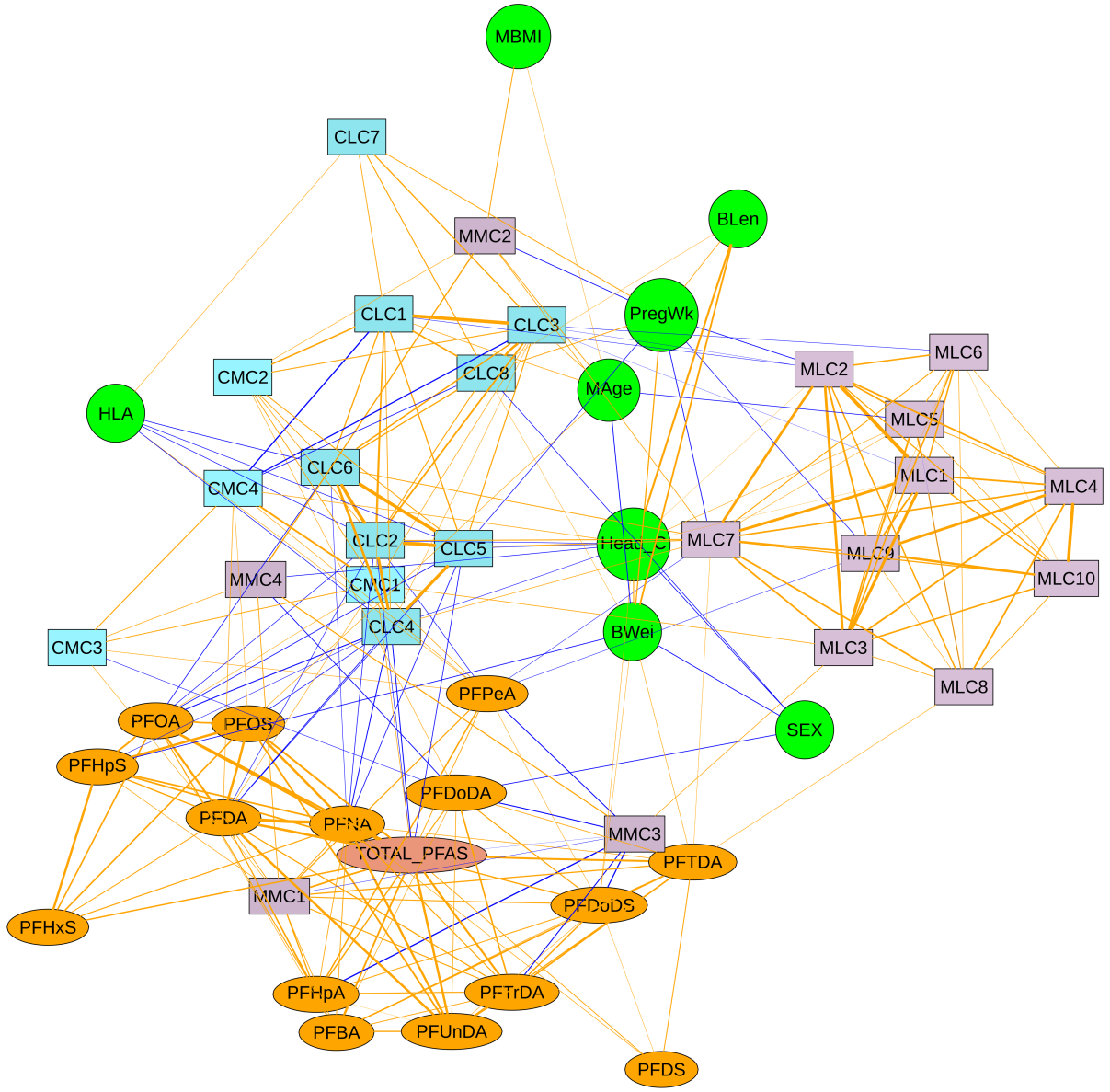

FIGURE 3

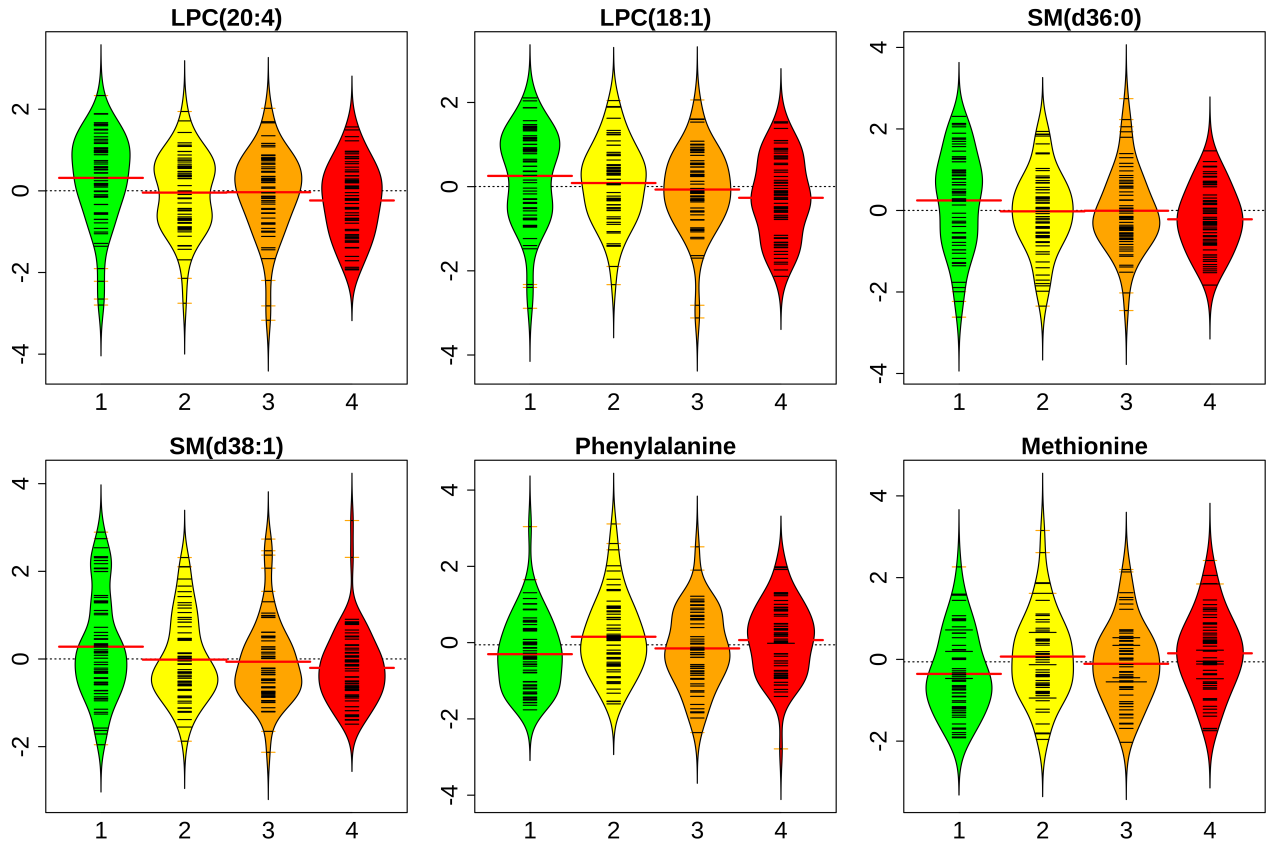

FIGURE 4

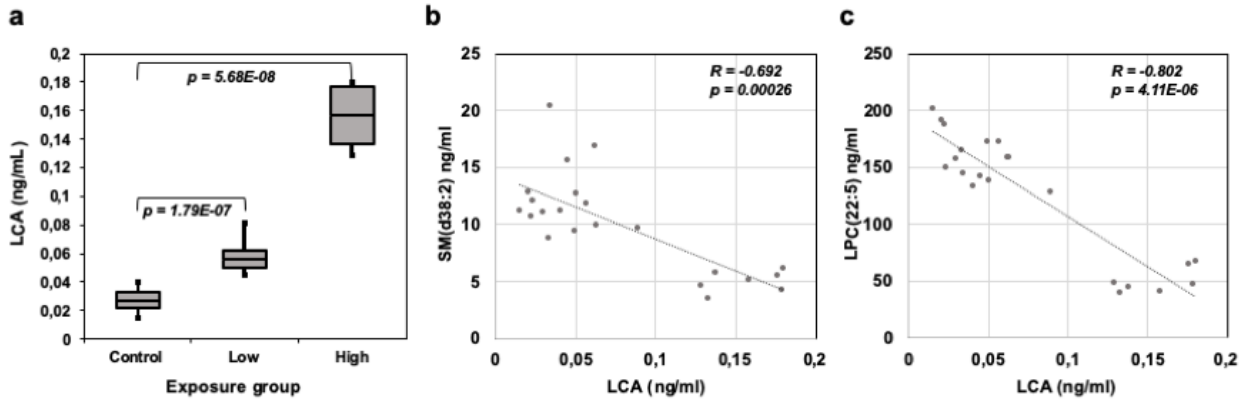

FIGURE 5

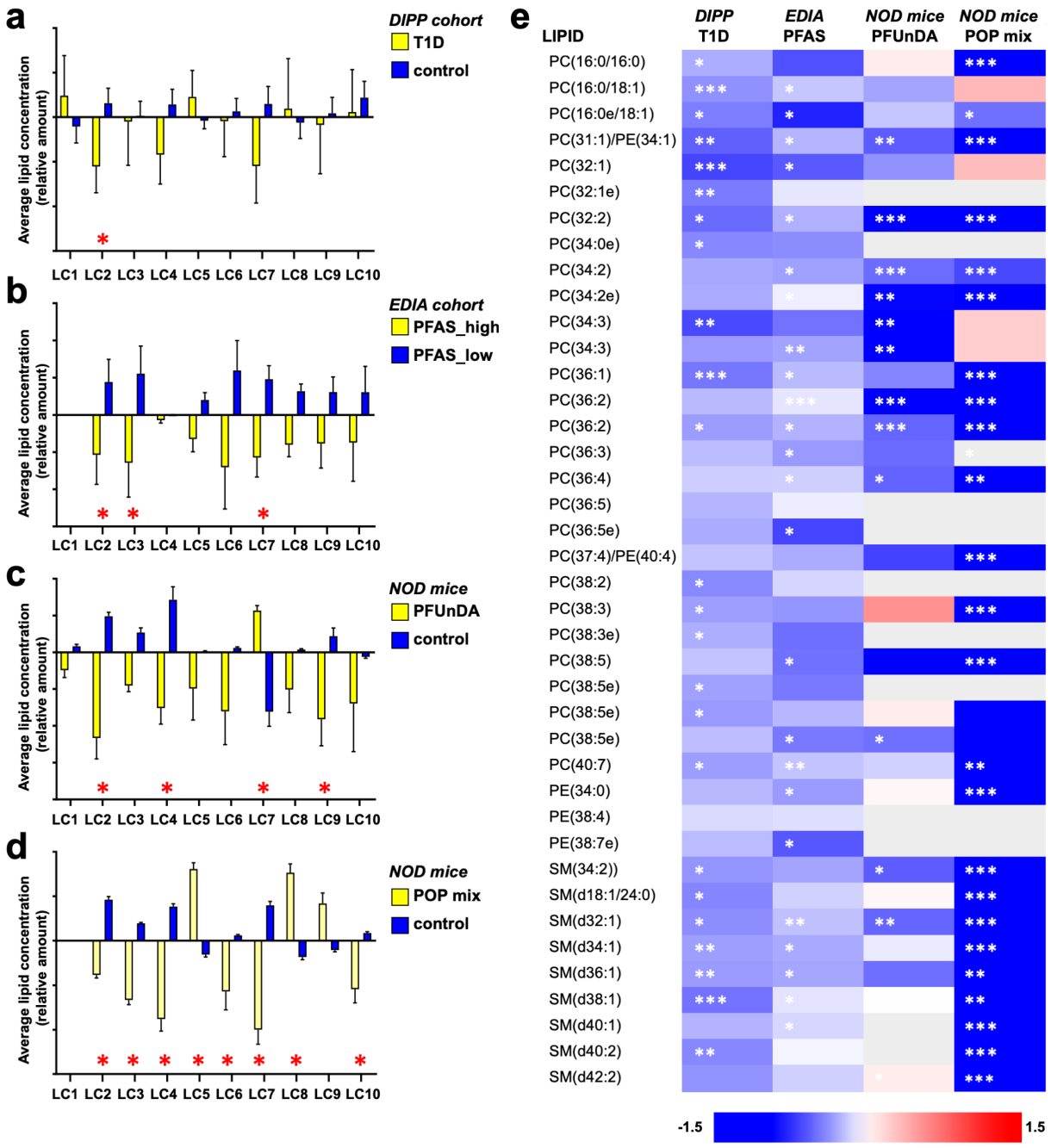

FIGURE 6