

OPEN

# Metabolite-related dietary patterns and the development of islet autoimmunity

Randi K. Johnson<sup>1</sup>, Lauren Vanderlinden<sup>2</sup>, Brian C. DeFelice<sup>3</sup>, Katerina Kechris<sup>2</sup>, Ulla Uusitalo<sup>4</sup>, Oliver Fiehn<sup>3,5</sup>, Marci Sontag<sup>1</sup>, Tessa Crume<sup>1</sup>, Andreas Beyerlein<sup>6,7</sup>, Åke Lernmark<sup>8</sup>, Jorma Toppari<sup>9,10</sup>, Anette-G. Ziegler<sup>7</sup>, Jin-Xiong She<sup>11</sup>, William Hagopian<sup>12</sup>, Marian Rewers<sup>13</sup>, Beena Akolkar<sup>14</sup>, Jeffrey Krischer<sup>4</sup>, Suvi M. Virtanen<sup>15,16,17</sup>, Jill M. Norris<sup>1</sup> & The TEDDY Study Group\*

The role of diet in type 1 diabetes development is poorly understood. Metabolites, which reflect dietary response, may help elucidate this role. We explored metabolomics and lipidomics differences between 352 cases of islet autoimmunity (IA) and controls in the TEDDY (The Environmental Determinants of Diabetes in the Young) study. We created dietary patterns reflecting pre-IA metabolite differences between groups and examined their association with IA. Secondary outcomes included IA cases positive for multiple autoantibodies (mAb+). The association of 853 plasma metabolites with outcomes was tested at seroconversion to IA, just prior to seroconversion, and during infancy. Key compounds in enriched metabolite sets were used to create dietary patterns reflecting metabolite composition, which were then tested for association with outcomes in the nested case-control subset and the full TEDDY cohort. Unsaturated phosphatidylcholines, sphingomyelins, phosphatidylethanolamines, glucosylceramides, and phospholipid ethers in infancy were inversely associated with mAb+ risk, while dicarboxylic acids were associated with an increased risk. An infancy dietary pattern representing higher levels of unsaturated phosphatidylcholines and phospholipid ethers, and lower sphingomyelins was protective for mAb+ in the nested case-control study only. Characterization of this high-risk infant metabolomics profile may help shape the future of early diagnosis or prevention efforts.

Type 1 diabetes affects over 500,000 children globally, making it one of the most common metabolic illnesses in children<sup>1</sup>. Autoimmune destruction of the insulin-producing beta cells in the pancreas results in hyperglycemia and lifelong insulin dependency. Genetic risk factors are well described and likely interact with non-genetic risk factors to influence disease progression, though exact pathogenesis remains unclear<sup>2</sup>. The appearance of autoantibodies can be detected as early as 3 months of age and defines the beginning of islet autoimmunity (IA), the preclinical stage of the disease<sup>3</sup>. Efforts to better characterize metabolic dysregulation around the time of

<sup>1</sup>Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, USA. <sup>2</sup>Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, USA. <sup>3</sup>UC Davis Genome Center—Metabolomics, University of California Davis, Davis, USA. <sup>4</sup>Health Informatics Institute, University of South Florida, Tampa, USA. <sup>5</sup>Department of Molecular and Cellular Biology, University of California Davis, Davis, USA. <sup>6</sup>Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany. <sup>7</sup>Institute of Diabetes Research, Helmholtz Zentrum München, and Klinikum rechts der Isar, Technische Universität München, and Forschergruppe Diabetes e.V., Neuherberg, Germany. <sup>8</sup>Department of Clinical Sciences, Lund University/CRC, Lund, Sweden. <sup>9</sup>Department of Pediatrics, Turku University Hospital, Turku, Finland. <sup>10</sup>Institute of Biomedicine, Research Centre for Integrated Physiology and Pharmacology, University of Turku, Turku, Finland. <sup>11</sup>Center for Biotechnology and Genomic Medicine, Augusta University, Augusta, USA. <sup>12</sup>Pacific Northwest Diabetes Institute, Seattle, USA. <sup>13</sup>Barbara Davis Center for Childhood Diabetes, University of Colorado Anschutz Medical Campus, Aurora, USA. <sup>14</sup>National Institutes of Diabetes and Digestive and Kidney Disorders, National Institutes of Health, Bethesda, USA. <sup>15</sup>National Institute for Health and Welfare, Tampere, Finland. <sup>16</sup>University of Tampere, Tampere, Finland. <sup>17</sup>Tampere University Hospital, Tampere, Finland. \*A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to J.M.N. (email: [Jill.norris@cuanschutz.edu](mailto:Jill.norris@cuanschutz.edu))

seroconversion and prior to the detection of autoantibodies may allow earlier identification of at-risk children and better understanding of the processes involved.

Metabolites reflect the interaction of numerous biological factors, including many that may influence the development of autoimmune diabetes, such as genetics, microbiome, and dietary intake. Metabolomics differences between IA cases and controls mostly have been found at the time of seroconversion, but are inconsistent across studies conducted in country-specific populations<sup>4–7</sup>. Previous country-specific studies used different laboratories to measure varying types (primary/polar<sup>5</sup>, lipids<sup>4,7–9</sup>, both<sup>6,10</sup>) and amounts of metabolomics features (from 106<sup>7</sup> to 540<sup>6</sup>), and conducted studies at varying ages and stages of the disease course (cord blood<sup>4,7,8</sup>, at seroconversion to IA<sup>6</sup>, longitudinally<sup>5,9,10</sup>). Given these methodological differences, it is unclear whether differences in study findings are due to technical artefacts, or whether they represent truly different associations by geography or other meaningful characteristic. The identification of generalizable metabolic profiles related to the development of early stages of the disease across several populations is important and may inform dietary interventions to prevent type 1 diabetes, which have so far proven unsuccessful<sup>11</sup>.

Traditional investigation of diet in the development of type 1 diabetes has examined effects of individual foods or food groups and nutrients such as cow's milk<sup>11–13</sup>, fatty acids<sup>14–18</sup>, or vitamin D<sup>19–22</sup>. However, these approaches do not account for the complexity of the diet—the effects of single nutrients and foods are often too small to identify, or too highly correlated to be separated from each other<sup>23</sup>. Examining combinations of foods and metabolites may better elucidate the role of diet in IA, as it can account for synergistic or antagonistic effects of foods or nutrients contained in the diet, and differences in how they are processed in the body.

We aimed to identify metabolite-related dietary patterns associated with IA in the multinational The Environmental Determinants of Diabetes in the Young (TEDDY) study. We conducted a metabolome- and lipidome-wide association study to better characterize plasma metabolites and lipids distinguishing cases and controls both at the time of the first autoantibody detection, and prior to its development. We created dietary patterns summarizing candidate metabolites identified pre-IA, and tested the longitudinal association of those metabolite-related dietary patterns with the development of IA.

## Methods

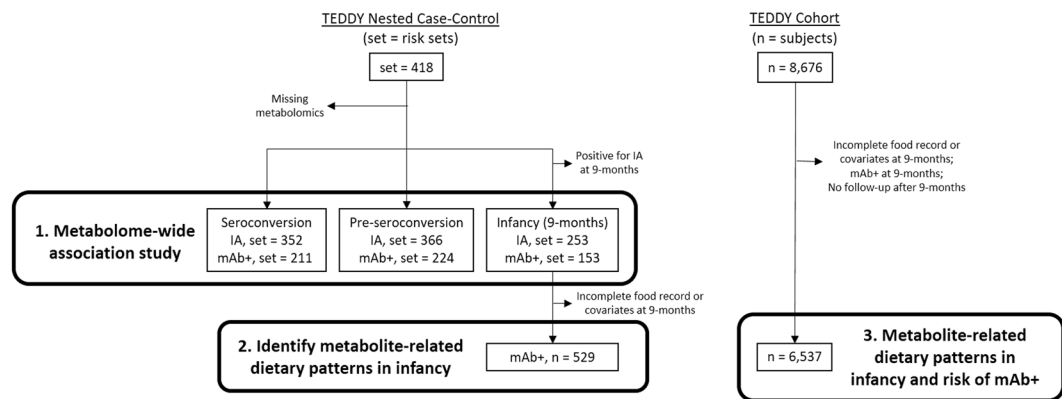
**TEDDY study design.** TEDDY is an international consortium that enrolled 8,676 newborn infants with a high- or moderate-risk class II HLA genotype between 2004 and 2010<sup>24</sup>. Participants are closely followed for the development of IA or type 1 diabetes, with study visits every three months from birth to age 48 months, and every three or six months thereafter depending on autoantibody status until the age of 15 years. Participating study centers include: Georgia/Florida, Colorado, and Washington in the U.S., and Finland, Sweden, and Germany in Europe. IA cases are defined by confirmed autoantibody positivity to either insulin (IAA), GAD (GADA), or IA-2 (IA-2A) on two consecutive study samples, the first of which defines the case's event age.

A nested case-control biomarker study was designed using risk set sampling to select three controls per IA case ( $n = 418$ ) that had developed in TEDDY as of May 2012. Eligible controls were autoantibody-negative at the case's event age, and further matched on clinical center, sex, and family history of type 1 diabetes as previously described<sup>25</sup>. Secondary outcomes included cases positive for IAA only or GADA only at IA event time. IA-2A was excluded as an outcome since very few cases developed IA2 as their first and only persistent confirmed autoantibody at IA case-time. Multiple autoantibody positivity (mAb+) was defined as any subject positive for more than one autoantibody at IA event time, or who developed more than one autoantibody during follow-up.

The study methods have been carried out in accordance with the approved guidelines by local Institutional Review or Ethics Boards, including: Colorado Multiple Institutional Review Board (#04-0361); Medical College of Georgia Human Assurance Committee (2004–2010)/Georgia Health Sciences University Human Assurance Committee (2011–2012)/Georgia Regents University Institutional Review Board (2013–2017)/Augusta University Institutional Review Board (2017-present) (#HAC 0405380); University of Florida Health Center Institutional Review Board (#IRB201600277); Washington State Institutional Review Board (2004–2012)/Western Institutional Review Board (2013–present) (#20130211); Ethics Committee of the Hospital District of Southwest Finland (#Dnro168/2004); Bayerischen Landesärztekammer (Bavarian Medical Association) Ethics Committee (#04089); and Regional Ethics Board in Lund, Section 2 (2004–2012)/Lund University Committee for Continuing Ethical Review (2013-present) (#217/2004). The study is monitored by an External Evaluation Committee formed by the National Institutes of Health. Written informed consents were obtained from a parent and/or legal guardian for all participating children. Data described in the manuscript and code book will be made available upon request from the NIDDK Central Repository at <https://www.niddkrepository.org/studies/teddy>.

**Metabolomics data pre-processing.** Metabolomics abundance measures (metabolites and lipids) were obtained for all cases and controls for each available study visit from birth until the case event time. Primary metabolites and complex lipids were quantified from citrate plasma using GC-TOF MS and CSH-QTOF MS data acquisition, respectively, at the NIH West Coast Metabolomics Center at the University of California, Davis. GC-TOF MS data were acquired as previously described<sup>26</sup>, with data processing and compound identification using the BinBase algorithm<sup>27</sup>, GC-TOF data were sum normalized followed by LOESS (locally weighted scatterplot smoothing) normalization. For complex lipids, samples were extracted by methyl-tert-butyl ether/methanol/water<sup>28</sup>, followed by chromatogram peak detection and alignments using Mass Profiler Professional (Agilent, Santa Clara, CA). Peaks detected in a minimum of 30% of samples were identified and quantification back-filled using the Fiehn laboratory's LipidBlast spectral library, as previously described<sup>29</sup>. LOESS followed by batch ratio (QC samples were used to adjust sample batch median to global study median) normalization was performed across all the samples to estimate and remove analytical variance.

Prior to transformation, data quality checks included evaluation at the metabolite- and sample-level (Supplemental Fig. 1). Metabolites that were not detected in more than 10% of samples (6 metabolites), or with



**Figure 1.** Data flow diagram summarizing selection and size of analysis population for all aims (numbered). From a nested case-control study in TEDDY, we tested the association of 853 metabolites with outcomes at the time of seroconversion to IA (sets = 352), the last sample prior to IA (sets = 366), and at infancy (sets = 253). We created dietary patterns explaining candidate mAb+ metabolites identified in infancy, when children (n = 529) were 9-months of age and autoantibody negative. All subjects with food records at 9-months in the full TEDDY cohort (n = 6,537) were scored on the dietary pattern, and the association with development of mAb+ tested. IA = islet autoimmunity, mAb+ = multiple autoantibody positive, sets = number of risk sets or matched strata, n = number of subjects.

a coefficient of variation greater than or equal to 100% (286 metabolites) were excluded from further analyses. Samples with missing or zero values in greater than 10% of metabolomics features (n = 5) or with values more extreme than 4 standard deviations above or below the mean in greater than 30% of metabolomics features (n = 6) were removed from analyses. A total of 853 metabolites and lipids and 11,556 samples passed the quality checks. All metabolites were transformed using Box-Cox transformation analysis, and scaled<sup>30</sup>.

**Dietary intake and food groupings.** Dietary assessment was carried out by 24-hour recall at the first clinic visit at 3–4.5 months of age, then by 3-day food record every 3 months until 12 months of age, and then every 6 months thereafter. TEDDY research staff provided detailed instruction and examples to families regarding completion of food records, as previously described<sup>31</sup>. From quantities of foods and dishes consumed, the amounts of energy and single foods contained therein were computed using in-house food record processing programs and food composition databases unique to each country<sup>32</sup>. The foods and dishes (e.g. wheat bread, apple-oat meal) consumed were quantified into main food groups (ie: cereals, fruits and berries, etc.) and subgroups (ie: wheat, rice, oats, citrus fruits, apple, berries, etc.) in grams per day (g/day) of intake. After quantification, the three food records were averaged to calculate the mean energy and food intake for each study subject on each study visit. Results of detailed harmonization studies of these country-specific food composition databases documented that the energy values and food subgroups used in this study were comparable across the TEDDY countries<sup>32</sup>.

For any food record where a subject was indicated as breast fed, we estimated the amount of breastmilk consumption using an algorithm developed by the Institute of Medicine<sup>33</sup>. First, we calculated the estimated energy requirement based on age and weight. The difference in the estimated energy requirement and the mean energy reported on the food record from food and formula was attributed to breastmilk. We calculated the amount (grams) of breastmilk consumed to achieve that energy intake using a conversion factor of energy density per 100 g, as follows: 65.3 kcal/100 g in Finland, 69 kcal/100 g in Germany, 68 kcal/100 g in Sweden, and 70 kcal/100 g in the U.S.

**Statistical analyses.** Statistical analyses are described by aim below. First, we identified plasma metabolites and lipids associated with IA at three different time points, using a metabolome-wide association approach in TEDDY's nested case-control study. Second, we created dietary patterns summarizing candidate metabolites identified in infancy. Finally, we tested the longitudinal association of infancy metabolite-related dietary patterns with development of mAb+ in the full TEDDY cohort. Figure 1 summarizes the population and data flow for all aims and analyses.

**Metabolites associated with IA.** Conditional logistic regression was used to calculate odds ratios (ORs) for the association of each transformed metabolite with the development of IA, adjusting for high-risk HLA genotype (DR3-DQA1\*05:01-DQB1\*02:01/DR4-DQA1\*03:01-DQB1\*03:02 versus all other), and age at blood draw. Some TEDDY subjects follow a long-distance protocol, in which blood is drawn and shipped to clinical centers before being processed for biomarker identification. Since plasma primary GC-TOF MS metabolic profiles are less stable with centrifugation delay<sup>34</sup>, we required an additional match for the long distance protocol between case and control samples.

Metabolomics analyses were run in three cross-sections. First, we examined metabolite differences between IA cases and controls at the first detection of autoantibody positivity (seroconversion), defined as the first of the two consecutive autoantibody positive visits for cases. Then, to identify metabolites and lipids that may differentiate IA cases and controls prior to the detection of autoantibodies (pre-seroconversion) we selected the most recent

IA-free visit for cases. Finally, since autoimmunity can begin very early in life and both metabolomics and dietary factors are strongly related to age, we identified metabolites distinguishing IA cases and controls prior to the appearance of autoantibodies in infancy. The “infancy” cross-section was defined as an IA-free visit at 9 months of age for cases. For controls, the visit corresponding to the case visit was selected for all cross-sections. Children positive for autoantibodies at 9-months ( $n = 48$  IA cases) and their matched controls were excluded from the infancy cross-section (Fig. 1).

We tested the association of each metabolite with the secondary outcomes described above: IAA, GADA, and mAb+. We considered  $p$ -value  $< 0.05$  significant since traditional approaches for multiple comparison correction may be too strict for the unusually highly correlated metabolomics data or inappropriate given the exploratory nature of our study aims<sup>35</sup>. SAS version 9.4 was used for these analyses.

We focused on pathway enrichment, given that metabolites may capture perturbations in many upstream biological systems thereby complicating interpretation of individual associations. ChemRICH forms non-overlapping groups of metabolites based on chemical similarity and ontology mapping<sup>36</sup>. It calculated a single  $p$ -value for each group, and identified the most significant metabolite in each group as the “key compound” (<http://chemrich.fiehnlab.ucdavis.edu/>). Inputs for the ChemRICH analyses included the nominal  $p$ -value and odds ratio from the individual conditional logistic regression models, and chemical structure information from well-characterized known metabolites and lipids ( $p = 315$ , see Supplemental Table 1).

*Metabolite-related dietary patterns preceding the appearance of autoantibodies.* Reduced rank regression (RRR) was used to identify dietary patterns reflecting metabolites associated with IA. RRR creates linear combinations of foods (dietary patterns) that explain the maximum covariation in a second set of intermediate response variables (metabolites)<sup>37</sup>, thereby capturing disease-related variation in the diet rather than general eating behaviors identified from other dietary pattern methods<sup>38</sup>. We focused dietary pattern analysis on the infancy cross-section, since it is prior to the beginning of the autoimmune process and all children were the same age. Conducting dietary pattern analyses with foods in young children of different ages could be problematic, since not all foods are able to be eaten at all ages.

Foods from the food record were combined into 43 subgroups based on nutrient content and culinary usage (Supplemental Table 2)<sup>32</sup>. Intake of various foods varies greatly at 9-months of age, leading to some foods having a large proportion of subjects with no reported intake. Therefore, we filtered out food subgroups with a high proportion of non-eaters, as a way to deal with inflated zero distributions, as is common in studies using RRR<sup>39</sup>. Food subgroups that were shown to be comparable across TEDDY countries (ie harmonized), and were eaten by at least 40% of subjects in infancy were included in the creation of dietary patterns. The resulting dietary patterns therefore reflect foods that are most commonly eaten at 9-months among TEDDY study participants. Food subgroups were standardized to the age-specific mean and standard deviation of all TEDDY food records for dietary pattern analyses. The key compound in each significantly enriched metabolite group (identified by ChemRICH) was used as RRR response variables.

The number of dietary patterns needed to best explain the variation in metabolites was selected using the van der Voet T2 statistic<sup>40</sup>. The loadings (or relative weights) of food groups on each dietary pattern and the partial correlation with metabolite response variables was used to interpret each dietary pattern.

*Metabolite-related dietary patterns and risk of IA.* Infancy metabolite-related dietary patterns were first tested in the nested case-control study using conditional logistic regression as described above. Then we applied them to the full TEDDY cohort at 9 months of age, using the food group loadings to generate one score per pattern for each subject with complete food records ( $n = 6,537$ ). The dietary pattern score is a linear combination of food intakes using the reported intake of each food weighted by the factor loadings<sup>37</sup>. The score indicates how similar the reported dietary intake of a subject is to the dietary pattern—with higher scores indicating a diet similar to the pattern, and lower scores indicating a diet dissimilar to the pattern.

Cox proportional-hazards models were used to test the association of metabolite-related dietary pattern scores at 9-months on risk of mAb+, adjusting for clinical center, high-risk HLA genotype, family history of type 1 diabetes, total energy intake, and sex. Adjustment factors used in multivariable models were selected based on standard adjustments used in the TEDDY study<sup>31,41</sup>. Time-to-event analyses were performed to evaluate whether metabolite-related dietary pattern scores were associated with mAb+ by the age of 6 years. Cases included those selected for the nested case-control study plus any additional cases that developed by January 2018, when the prospective data collection was cut for analyses. Given that risk factors for IA may differ by age, we restricted follow up to 6 years in the TEDDY cohort to ensure the cohort analysis represented a similarly-aged case-population as the nested case-control study. For consistency with the nested case-control study, the time-to-event was defined as the time from birth to the appearance of the first persistent confirmed autoantibody among IA cases who developed a second persistent confirmed autoantibody at any point. Subjects without complete covariate information, those developing mAb+ at 9-months ( $n = 53$ ), or who had no follow-up after 9-months ( $n = 243$ ) were excluded from survival analysis (Fig. 1).

## Results

**Metabolic dysregulation apparent at seroconversion and infancy.** From the nested case-control study, there were 352 matched sets with metabolomics measures for 1 case and at least 1 control at seroconversion (mean (SD) case-age = 722 (446) days), 366 sets at pre-seroconversion (mean (SD) = 625 (412) days), and 253 sets at the 9-month infancy visit (mean (SD) = 283 (14) days) (Table 1). For secondary outcomes, 49% of IA cases were positive for IAA, 32% for GADA, and 60% for mAb+. The distribution of secondary outcomes was consistent across the seroconversion, pre-seroconversion, and 9-month infancy cross-sections. The majority of IA cases were from Sweden (32%) and Finland (28%). IA cases were 55% male, 22% had a first degree relative, and

Cross-section	IA		IAA*			GADA*			mAb+*		
	n	Case-age <sup>†</sup> , mean (SD)	n	% of IA cases	Case-age, mean (SD)	n	% of IA cases	Case-age, mean (SD)	n	% of IA cases	Case-age, mean (SD)
Seroconversion	352	722 (446)	171	48.6	586 (370)	113	32.1	888 (509)	211	59.9	655 (365)
Pre-Seroconversion	366	625 (412)	180	49.2	505 (366)	116	31.7	786 (445)	224	61.2	541 (346)
Infancy 9-months	253	283 (14)	114	45.1	283 (14)	83	32.8	284 (16)	153	60.5	282 (14)

**Table 1.** Description of matched sets (1 case and 1, 2 or 3 controls) for metabolomics analyses by outcome and cross-section. \*Secondary outcomes defined as IAA or GADA as first-appearing and only autoantibody at IA case-time (mutually exclusive), while mAb+ indicates IA case developed more than 1 persistent confirmed Ab at any point during follow-up. <sup>†</sup>Age at the time of metabolomics blood draw, in days. n = Number of matched sets (each set has 1 case and 1, 2, or 3 controls).

12% had their seroconversion blood-draw following TEDDY's long distance protocol (Supplemental Table 3). The distribution of matched sets by matching factors (clinical center, sex, first-degree relative status, and long distance protocol) was similar in the secondary outcomes compared to primary IA. Approximately 9% of the pre-seroconversion case samples (n = 34) occurred during the 9-month infancy visit.

Conditional logistic regression results from the metabolome-wide association study indicated metabolic dysregulation in cases compared to matched controls at seroconversion and during infancy (Supplemental Fig. 2). More metabolites and lipids were different by case status when restricting to the mAb+ outcome (p = 130, 15%) compared to the IA outcome (p = 64, 7.5%). There were few metabolites associated with the secondary outcomes IAA first and GADA first in any cross-section. Therefore, we focused the metabolomics set enrichment analyses on the mAb+ outcome, which represented 60% of IA cases and had the largest signal of metabolomics differences between cases and controls. ChemRICH identified seven groups of chemically similar metabolites that were significantly different among mAb+ cases and controls at seroconversion, one group at pre-seroconversion, and six groups in infancy (Fig. 2, Supplemental Table 4).

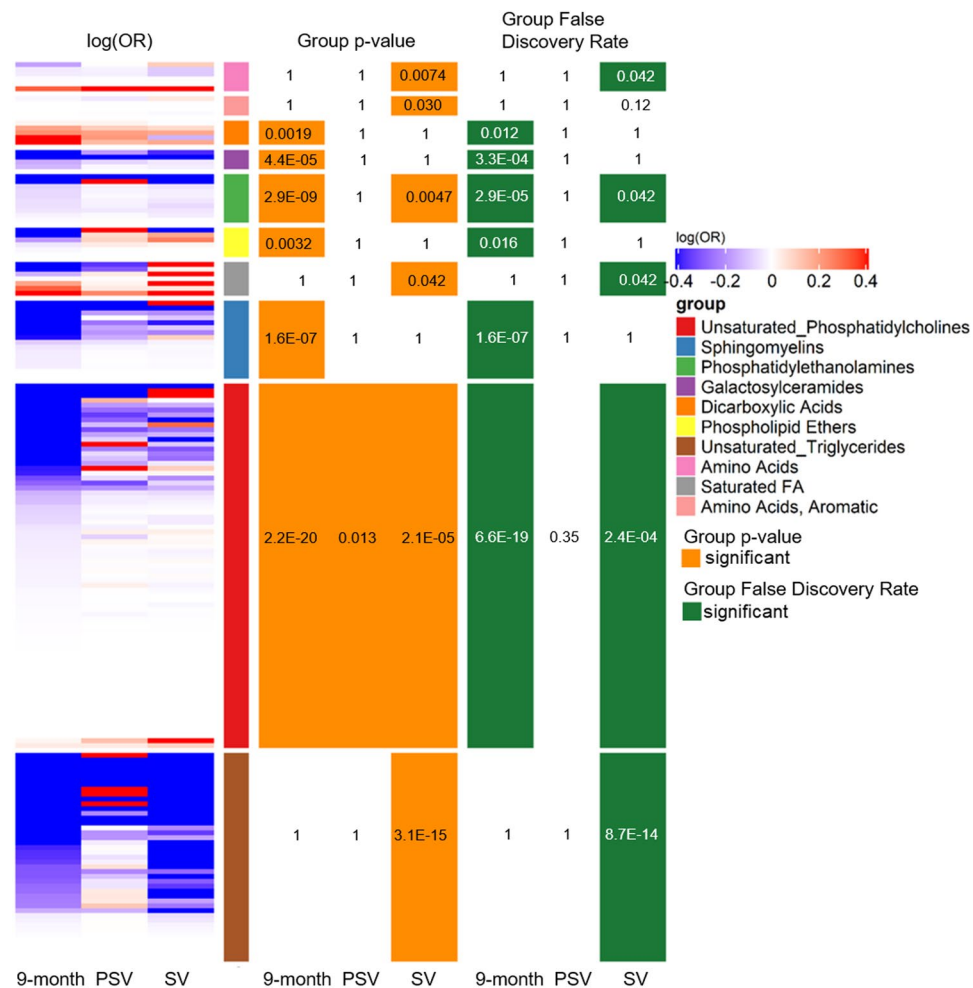
Of the metabolite groups identified as different between mAb+ cases and controls, only unsaturated phosphatidylcholines (PC) were consistently dysregulated in all three analyses (p-value for group in seroconversion =  $2.1 \times 10^{-5}$ , pre-seroconversion = 0.013, infancy =  $2.2 \times 10^{-20}$ ), with the majority of the individual metabolites being lower in mAb+ cases compared with controls (OR < 1) (Fig. 2). Similarly, phosphatidylethanolamines (PE) were lower in mAb+ cases (OR < 1) at both seroconversion (p-value = 0.0047) and in infancy (p-value =  $2.9 \times 10^{-6}$ ).

Other than PCs and PEs, distinct metabolite groups distinguished mAb+ cases from their controls at the time of seroconversion to primary IA compared to during infancy prior to the appearance of any autoantibodies. At seroconversion, mAb+ cases had lower levels of unsaturated triglycerides (p-value =  $3.1 \times 10^{-15}$ ), amino acids (p-value = 0.0074), diglycerides (p-value = 0.014), and aromatic amino acids (p-value = 0.03), and higher levels of saturated fatty acids (p-value = 0.0074). In infancy, other phospholipids were significantly protective for mAb+ (majority of OR < 1), including sphingomyelins (SM, p-value =  $1.1 \times 10^{-8}$ ) and phospholipid ethers (EtherPL, p-value = 0.0032), along with the glucosylceramides (GlcCer, p-value =  $4.4 \times 10^{-5}$ ). Three dicarboxylic acids were significantly higher in mAb+ cases compared to controls in infancy (OR > 1, p-group = 0.0019).

**Infant metabolite-related dietary patterns and risk of mAb+.** From each of the six groups identified by ChemRICH in infancy, we used the key metabolite (most significant one) as a response variable in dietary pattern analyses, including: PC (34:3), SM (d41:2) A, PE (34:2), GlcCer (d41:1), adipic acid, and PC (p-32:0) or PC (o-32:1) (EtherPL).

Reduced rank regression identified three dietary patterns that explained 8% of the variation in metabolites and 29.3% of the food variation. Food groups factor loadings and metabolite variable weights used to interpret each dietary pattern are shown in Fig. 3. More extreme factor loadings or variable weights indicate the food or metabolite was influential in the dietary pattern. Infants scoring high on Dietary Pattern 1 ate more non-gluten containing cereals, onions, vegetable oils, and fat-free milk (positive factor loadings), and less breast milk (negative factor loadings). This diet corresponded to higher levels of PE (34:2), as indicated by the higher variable weight. Infants scoring high on Dietary Pattern 2 ate diets with higher saturated fats, fat-free milk, poultry, and infant formula, and lower in potatoes and vegetable oils. This diet corresponded to higher levels of SM (d41:2) A, GlcCer (d41:1), and PC (p-32:0) or PC (o-32:1). Finally, 9-month infants scoring high on Dietary Pattern 3 ate diets higher in breast milk, red meat, potatoes, and cereals, and lower in processed fruits, legumes, and infant formula. High scores on Dietary Pattern 3 corresponded to higher levels of PC (34:3) and PC (p-32:0) or PC (o-32:1), and lower levels of SM (d41:2) A. The correlation between metabolites and dietary patterns followed similar patterns (Supplemental Table 5).

Dietary patterns generated from metabolites and food intake in the nested case-control study were applied to the full cohort to generate one metabolite-related dietary pattern score for each dietary pattern on all 9-month diet records. Subjects developing mAb+ by age 6 years were more likely to have a first-degree relative (FDR) with type 1 diabetes and to have high-risk HLA-DR3/4 genotypes (Supplemental Table 6). Dietary Pattern 3 was significantly protectively associated with mAb+ in the nested case-control study (OR = 0.67, 95%CI = 0.48–0.94, Table 2). However, there was no association seen in time-to-event analyses applied to the whole cohort and adjusted for clinical center, sex, HLA-DR3/4, and FDR (HR = 0.98, 95%CI = 0.83–1.16, Table 2). No other 9-month metabolite-related dietary patterns were associated with development of mAb+ in the TEDDY cohort. Results did not change in a sensitivity analysis in which we adjusted for an additional 17 covariates that TEDDY



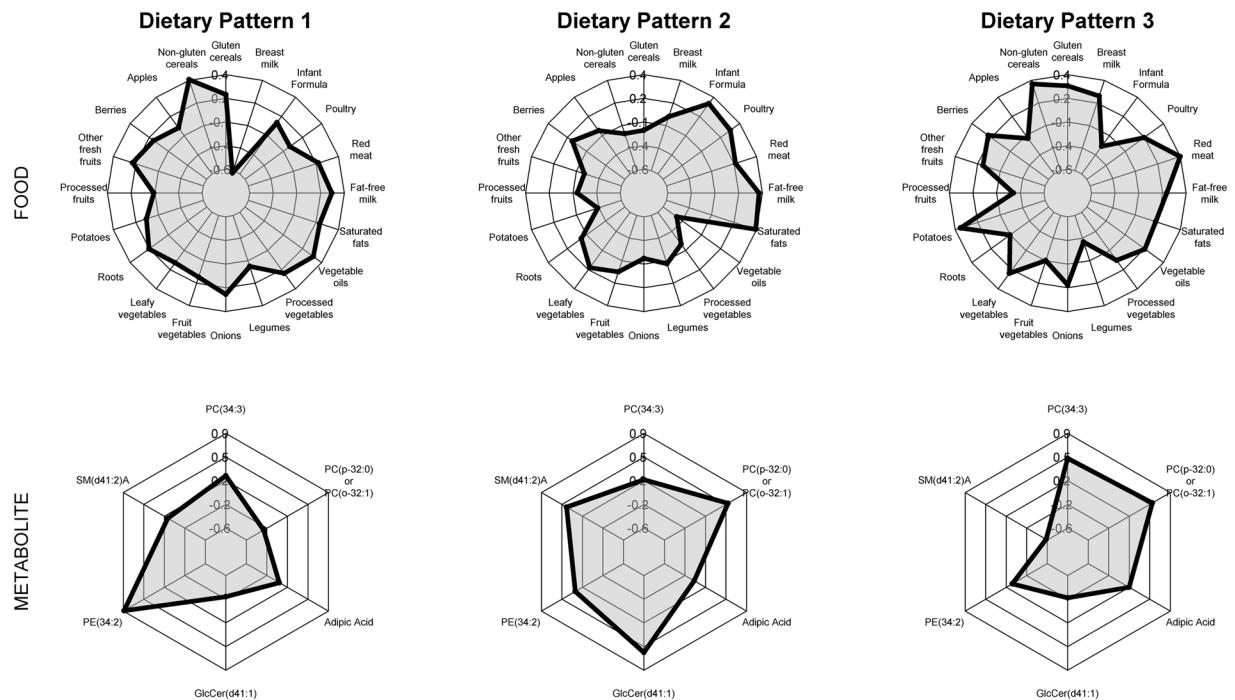
**Figure 2.** Chemically similar metabolite sets identified as significantly associated with mAb+ by ChemRICH. Each row is an individual metabolite, grouped by ChemRICH set and sorted by log(OR) within each set. Log(OR) > 0 (red) indicates a positive association between metabolite and mAb+, whereas log(OR) < 0 (blue) indicates an inverse association between metabolite and mAb+. Phosphatidylcholines were significantly lower in cases compared to controls in infancy (9-month), just prior to seroconversion (PSV), and at seroconversion to primary IA (SV). Other phospholipids were significantly lower in cases only in infancy, while other metabolite groups, such as unsaturated triglycerides and amino acids, distinguished cases and controls at seroconversion. Metabolite groups identified as significant (group p-value < 0.05) in any cross-section are shown, along with the corresponding adjusted p-value for the group (group false discovery rate).

has identified as associated with development of IA (race-ethnicity, maternal education, maternal age, introduction of probiotics before 28 days, introduction of probiotics at or after 28 days, weight for age z-score at 12 months, and number of minor alleles for rs2476601, rs2816316, rs11711054, rs10517086, rs4948088, rs1004446, rs7111341, rs2292239, rs3184504, rs3825932, rs12708716) (data not shown).

## Discussion

We identified dysregulated metabolism at the onset of and preceding stage 1 diabetes (mAb+) in a multi-national, prospective type 1 diabetes study. PC and PE metabolite groups were consistently decreased in mAb+ cases compared to controls both prior to and at the time of seroconversion. Unsaturated triglycerides and amino acid groups were lower among mAb+ cases at seroconversion only, whereas SM, GlcCer, and EtherPL lipids were lower among mAb+ cases in infancy only. While an infancy dietary pattern explaining choline- and sphingosine-containing lipids was associated with mAb+ in the nested case-control study, this association was not observed in the full TEDDY cohort.

Dicarboxylic acids were the only metabolite group we found associated with increased risk of mAb+ in infancy. Adipic acid was the key compound of the dicarboxylic acids group. Other dicarboxylic acids associated with increased mAb+ risk included the tricarboxylic acid (TCA) cycle intermediaries succinic acid and malic acid. While metabolomics studies in Norway and Germany did not identify TCA cycle metabolites<sup>5,6</sup>, a previous study in Finnish children found both succinic acid and glutamic acid were increased in type 1 diabetes cases 0–9 months prior to autoantibody appearance<sup>10</sup>. We were not able to examine glutamic acid as its measurement was



**Figure 3.** Food group loadings and metabolite weights for metabolite-related dietary patterns. In total, the three dietary patterns explained 8% of metabolite variation and 29.3% of food variation. For food, the radial axis indicates the loading on each dietary pattern (Range:  $-0.6$  to  $0.4$ ), and is used to interpret which combinations of foods are influential in the dietary pattern. Similarly, the metabolite radial axis indicates the weight of each metabolite on each dietary pattern (Range:  $-0.6$  to  $0.9$ ), indicating which combination of metabolites are explained by each dietary pattern. For example, subjects scoring high on dietary pattern 1 had diets higher in non-gluten containing cereals, onions, vegetable oils, and fat-free milk, and lower in breast milk. This diet corresponded to higher levels of PE (34:2).

Metabolite-related dietary patterns	Nested Case-Control* n = 147 mAb+ cases		Cohort <sup>†</sup> n = 300 mAb+ cases by 6 years		
	OR	95%CI	HR	95%CI	
1	0.85	0.68 1.05	0.95	0.83	1.08
2	0.81	0.61 1.08	0.89	0.78	1.02
3	0.67	0.48 0.96	0.98	0.83	1.16

**Table 2.** Dietary patterns at 9-months of age associated with risk of mAb+ in TEDDY. \*Conditional logistic regression models adjusted for age at metabolomics blood draw and total energy. <sup>†</sup>Survival models adjusted for clinical center, sex, FDR, total energy, and HLA DR3/4. OR = Odds Ratio, CI = confidence interval, HR = Hazard Ratio.

inhibited by the use of citrate tubes for plasma collection and storage in TEDDY. Through their regulation of demethylase activity, succinic acid and other TCA cycle intermediaries may be important regulators of DNA and histone methylation<sup>42</sup>, which may have some links to type 1 diabetes pathogenesis<sup>43</sup>.

The remainder of metabolite groups distinguishing mAb+ cases from controls in infancy belonged to lipid classes, some of which have been inconsistently associated with type 1 diabetes endpoints. We identified phospholipid dysregulation of PC and PE metabolites prior to mAb+ in infancy and at the time of seroconversion. While Oresic *et al.*<sup>10</sup> similarly found lower phosphatidylcholines in children who later developed type 1 diabetes, Pflueger *et al.* found higher levels of triglycerides and PUFA-containing phosphatidylcholines in autoantibody-positive children<sup>6</sup>. Unsurprisingly, the direction of association is different before and after the appearance of autoantibodies, as has been reported for other risk factors for type 1 diabetes, such as erythrocyte membrane fatty acid levels<sup>15,17</sup> or diabetes susceptibility genes<sup>44</sup>.

Sphingolipid metabolism plays a role in diabetic pathologies, including regulating beta-cell apoptosis, proinsulin and insulin folding in the endoplasmic reticulum, and cytokine secretion<sup>45</sup>. The evidence supporting this connection has been recently extended from animal models into human islet cells<sup>46</sup>. We identified two sphingolipid groups as significantly lower in infancy for mAb+ cases versus controls, including the SM group, which were previously identified in type 1 diabetes metabolomics studies<sup>9,10</sup>, and the GlcCer group. As a whole, sphingolipids have been characterized as both pro- and anti-inflammatory. Endogenous sphingolipids are metabolically

involved in T-cell regulation, autoimmunity, and inflammation<sup>47</sup>, yet consumption of dietary sphingolipids have been linked to anti-inflammatory responses<sup>48</sup>. The protective effects of dietary sphingolipids may operate via changes in gut microbiota or by activating other cofactors such as peroxisome proliferator-activated receptor  $\gamma$  expression<sup>49</sup>, both of which have been implicated in type 1 diabetes<sup>50</sup>. While outside the scope of this study, targeted characterization of the relationship between sphingolipid dietary intake and metabolite levels in future research might help to disentangle reported contrasting effects, which likely depend on other factors such as specific sphingolipid structure and existing metabolic state.

We identified choline-containing lipid groups (PC, SM, EtherPL) as protective for development of mAb+ at 9 months of age, consistent with previous studies conducted at birth and 3 months of age<sup>7,9,10</sup>. Choline is important for rapid growth and development in infancy, as a constituent of phospholipid cellular membranes. Additionally, it may play a role in insulin resistance or energy metabolism, perhaps through its role as a methyl donor for epigenetic changes<sup>51</sup>.

A metabolite-related dietary pattern reflecting choline-containing foods and metabolites was protective for mAb+ in the nested case-control study; however, no dietary pattern at 9-months was associated with development of mAb+ by age six years in the full TEDDY cohort. There are several factors that could contribute to the lack of dietary pattern association found. First, untargeted metabolomics and 3-day food records may not be measured precisely enough to successfully identify disease-related dietary patterns at such a young age where variability in both is large. Second, using the most significant metabolite in each group may not be the best choice of response variable, which is the variable that determines the ability of reduced rank regression to capture disease-related variation in the diet<sup>52</sup>. Metabolites may reflect other environmental factors, such as medication or microbiome, and therefore be poorly correlated with dietary intake. Metabolomics measures were further limited because they were quantified from non-fasting samples, which has been shown to differentially impact serum metabolic profiles related to dietary factors<sup>53</sup>.

We identified metabolic dysregulation prior to the detection of autoantibodies that distinguished children whose lifetime risk for symptomatic (Stage 3) type 1 diabetes approaches 100%<sup>54</sup>. Metabolomics differences were more apparent when comparing the high-risk mAb+ group to controls than comparing all cases of IA to controls. Few differences were identified by the type of first-appearing autoantibody. This metabolomics discovery was more comprehensive and generalizes to a broader population than previous studies. However, our exploratory approach using nominal p-value cutoffs may necessitate replication of these findings in another study. While novel application of dietary patterns summarizing candidate metabolites did not successfully extend outside of the nested case-control study, the approach may show promise for future work with targeted measurement of disease-related metabolites. Application of these methods that account for complex dietary intake are particularly important in type 1 diabetes research, since components of dietary intake are among the leading hypothesized environmental factors that act on high-risk genetic background to cause type 1 diabetes.

In the TEDDY study, higher levels of dicarboxylic acids and lower level PCs, SMs, PEs, EtherPLs, and GlcCers at 9-months of age was associated with increased risk of mAb+. Characterization of this profile may help shape the future of early diagnosis or prevention efforts.

## Data Availability

Data described in the manuscript and code book will be made available upon request from the NIDDK Central Repository at <https://www.niddkrepository.org/studies/teddy>.

## References

- International Diabetes Federation. *IDF diabetes atlas*. (International Diabetes Federation, 2015).
- Katsarou, A. *et al.* Type 1 diabetes mellitus. *Nat. Rev. Dis. Primer* **3**, 17016 (2017).
- Krischer, J. P. *et al.* The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia* **58**, 980–987 (2015).
- Orešič, M. *et al.* Cord Serum Lipidome in Prediction of Islet Autoimmunity and Type 1 Diabetes. *Diabetes* **62**, 3268–3274 (2013).
- Jørgenrud, B. *et al.* Longitudinal plasma metabolic profiles, infant feeding, and islet autoimmunity in the MIDIA study. *Pediatr. Diabetes* **18**, 111–119 (2017).
- Pflueger, M. *et al.* Age- and Islet Autoimmunity-Associated Differences in Amino Acid and Lipid Metabolites in Children at Risk for Type 1 Diabetes. *Diabetes* **60**, 2740–2747 (2011).
- Torre, D. L. *et al.* Decreased Cord-Blood Phospholipids in Young Age-at-Onset Type 1 Diabetes. *Diabetes* **62**, 3951–3956 (2013).
- Lamichhane, S. *et al.* Cord-Blood Lipidome in Progression to Islet Autoimmunity and Type 1 Diabetes. *Biomolecules* **9**, 33 (2019).
- 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).
- 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).
- Knip, M. *et al.* Effect of Hydrolyzed Infant Formula vs Conventional Formula on Risk of Type 1 Diabetes: The TRIGR Randomized Clinical Trial. *JAMA* **319**, 38–48 (2018).
- Lamb, M. M. *et al.* The effect of childhood cow's milk intake and HLA-DR genotype on risk of islet autoimmunity and type 1 diabetes: The Diabetes Autoimmunity Study in the Young. *Pediatr. Diabetes* **16**, 31–38 (2015).
- Virtanen, S. M. *et al.* Food consumption and advanced  $\beta$  cell autoimmunity in young children with HLA-conferred susceptibility to type 1 diabetes: a nested case-control design. *Am. J. Clin. Nutr.* **95**, 471–478 (2012).
- Norris, J. M. *et al.* Omega-3 polyunsaturated fatty acid intake and islet autoimmunity in children at increased risk for type 1 diabetes. *JAMA* **298**, 1420–1428 (2007).
- Miller, M. R. *et al.* Erythrocyte membrane omega-3 fatty acid levels and omega-3 fatty acid intake are not associated with conversion to type 1 diabetes in children with islet autoimmunity: The Diabetes Autoimmunity Study in the Young (DAISY). *Pediatr. Diabetes* **12**, 669–675 (2011).
- Virtanen, S. M. *et al.* Serum fatty acids and risk of advanced  $\beta$ -cell autoimmunity: a nested case-control study among children with HLA-conferred susceptibility to type 1 diabetes. *Eur. J. Clin. Nutr.* **64**, 792–799 (2010).
- Norris, J. M. *et al.* Erythrocyte membrane docosapentaenoic acid levels are associated with islet autoimmunity: the Diabetes Autoimmunity Study in the Young. *Diabetologia* **57**, 295–304 (2014).



18. Niinistö, S. *et al.* Fatty acid status in infancy is associated with the risk of type 1 diabetes-associated autoimmunity. *Diabetologia* **60**, 1223–1233 (2017).
19. Hyppönen, E., Läärä, E., Reunanen, A., Järvelin, M. R. & Virtanen, S. M. Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. *Lancet Lond. Engl.* **358**, 1500–1503 (2001).
20. Simpson, M. *et al.* No association of vitamin D intake or 25-hydroxyvitamin D levels in childhood with risk of islet autoimmunity and type 1 diabetes: the Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia* **54**, 2779–2788 (2011).
21. Mäkinen, M. *et al.* Serum 25-Hydroxyvitamin D Concentrations in Children Progressing to Autoimmunity and Clinical Type 1 Diabetes. *J. Clin. Endocrinol. Metab.* **101**, 723–729 (2016).
22. Norris, J. M. *et al.* Plasma 25-Hydroxyvitamin D Concentration and Risk of Islet Autoimmunity. *Diabetes* **67**, 146–154 (2018).
23. Willett, W. *Nutritional Epidemiology*. (OUP USA, 2012).
24. The TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. *Pediatr. Diabetes* **8**, 286–298 (2007).
25. Lee, H.-S. *et al.* Biomarker discovery study design for type 1 diabetes in The Environmental Determinants of Diabetes in the Young (TEDDY) study. *Diabetes Metab. Res. Rev.* **30**, 424–434 (2014).
26. Fiehn, O. Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr. Protoc. Mol. Biol.* **114**, 30.4.1–30.4.32 (2016).
27. Kind, T. *et al.* FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* **81**, 10038–10048 (2009).
28. Cajka, T., Smilowitz, J. T. & Fiehn, O. Validating Quantitative Untargeted Lipidomics Across Nine Liquid Chromatography-High-Resolution Mass Spectrometry Platforms. *Anal. Chem.* **89**, 12360–12368 (2017).
29. Dunn, W. B. *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **6**, 1060–1083 (2011).
30. Box, G. E. & Cox, D. R. An analysis of transformations. *J. R. Stat. Soc. Ser. B Methodol.* 211–252 (1964).
31. Beyerlein, A. *et al.* Dietary intake of soluble fiber and risk of islet autoimmunity by 5 y of age: results from the TEDDY study. *Am. J. Clin. Nutr.* **102**, 345–352 (2015).
32. Joslowski, G. *et al.* Development of a harmonized food grouping system for between-country comparisons in the TEDDY Study. *J. Food Compos. Anal.* **63**, 79–88 (2017).
33. Trumbo, P., Schlicker, S., Yates, A. A., Poos, M. & Food and Nutrition Board of the Institute of Medicine, The National Academies. Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein and amino acids. *J. Am. Diet. Assoc.* **102**, 1621–1630 (2002).
34. Malm, L., Tybring, G., Moritz, T., Landin, B. & Galli, J. Metabolomic Quality Assessment of EDTA Plasma and Serum Samples. *Biopreservation Biobanking* **14**, 416–423 (2016).
35. Bender, R. & Lange, S. Adjusting for multiple testing—when and how? *J. Clin. Epidemiol.* **54**, 343–349 (2001).
36. Barupal, D. K. & Fiehn, O. Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Sci. Rep.* **7**, 14567 (2017).
37. Hoffmann, K., Schulze, M. B., Schienkiewitz, A., Nöthlings, U. & Boeing, H. Application of a New Statistical Method to Derive Dietary Patterns in Nutritional Epidemiology. *Am. J. Epidemiol.* **159**, 935–944 (2004).
38. DiBello, J. R. *et al.* Comparison of 3 Methods for Identifying Dietary Patterns Associated With Risk of Disease. *Am. J. Epidemiol.* **168**, 1433–1443 (2008).
39. Schulze, M. B. *et al.* Dietary pattern, inflammation, and incidence of type 2 diabetes in women. *Am. J. Clin. Nutr.* **82**, 675–684 (2005).
40. van der Voet, H. Comparing the predictive accuracy of models using a simple randomization test. *Chemom. Intell. Lab. Syst. Syst.* **25**, 313–323 (1994).
41. Uusitalo, U. *et al.* Early Infant Diet and Islet Autoimmunity in the TEDDY Study. *Diabetes Care* **41**, 522–530 (2018).
42. Salminen, A., Kaarniranta, K., Hiltunen, M. & Kauppinen, A. Krebs cycle dysfunction shapes epigenetic landscape of chromatin: novel insights into mitochondrial regulation of aging process. *Cell. Signal.* **26**, 1598–1603 (2014).
43. Paul, D. S. *et al.* Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. *Nat. Commun.* **7**, ncomms13555 (2016).
44. Krischer, J. P. *et al.* The Influence of Type 1 Diabetes Genetic Susceptibility Regions, Age, Sex, and Family History on the Progression From Multiple Autoantibodies to Type 1 Diabetes: A TEDDY Study Report. *Diabetes* **66**, 3122–3129 (2017).
45. Boslem, E., Meikle, P. J. & Biden, T. J. Roles of ceramide and sphingolipids in pancreatic  $\beta$ -cell function and dysfunction. *Islets* **4**, 177–187 (2012).
46. Holm, L. J. *et al.* Abnormal islet sphingolipid metabolism in type 1 diabetes. *Diabetologia* **61**, 1650–1661 (2018).
47. Maceyka, M. & Spiegel, S. Sphingolipid metabolites in inflammatory disease. *Nature* **510**, 58–67 (2014).
48. Norris, G. H. & Blesso, C. N. Dietary and Endogenous Sphingolipid Metabolism in Chronic Inflammation. *Nutrients* **9** (2017).
49. Mazzei, J. C. *et al.* Suppression of intestinal inflammation and inflammation-driven colon cancer in mice by dietary sphingomyelin: importance of peroxisome proliferator-activated receptor  $\gamma$  expression. *J. Nutr. Biochem.* **22**, 1160–1171 (2011).
50. Pörksen, S. *et al.* Variation within the PPAR $\gamma$  gene is associated with residual  $\beta$ -cell function and glycemic control in children and adolescents during the first year of clinical type 1 diabetes. *Pediatr. Diabetes* **9**, 297–302 (2008).
51. Zeisel, S. H. Metabolic crosstalk between choline/1-carbon metabolism and energy homeostasis. *Clin. Chem. Lab. Med. CCLM FESCC* **51**, 467–475 (2013).
52. Weikert, C. & Schulze, M. B. Evaluating dietary patterns: the role of reduced rank regression. *Curr. Opin. Clin. Nutr. Metab. Care* **19**, 341–346 (2016).
53. Sedlmeier, A. *et al.* The human metabolic profile reflects macro- and micronutrient intake distinctly according to fasting time. *Sci. Rep.* **8**, 12262 (2018).
54. Ziegler, A. G. *et al.* Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA* **309**, 2473–2479 (2013).

## Acknowledgements

Thank you to the participants and families of the TEDDY study, whose continued commitment make such research possible. The TEDDY Study is funded by U01 DK63829, U01 DK63861, U01 DK63821, U01 DK63865, U01 DK63863, U01 DK63836, U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 DK63821, UC4 DK63865, UC4 DK63863, UC4 DK63836, UC4 DK95300, UC4 DK100238, UC4 DK106955, UC4 DK112243, UC4 DK117483, and Contract No. HHSN267200700014C from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Child Health and Human Development (NICHD), National Institute of Environmental Health Sciences (NIEHS), Centers for Disease Control and Prevention (CDC), and JDRE. This work supported in part by the NIH/NCATS Clinical and Translational Science Awards to the University of Florida (UL1 TR000064) and the University of Colorado (UL1 TR001082). RKJ, LV, KK, OF, MR, and JMN were also funded by R01-DK104351.

### Author Contributions

R.K.J. led the design of the study, the statistical analysis and interpretation of data, and drafted the manuscript. L.V. analyzed and interpreted the metabolomics data, created figures, and critically reviewed the manuscript for intellectual content. B.C.D. led acquisition of the metabolomics data and critically reviewed the manuscript for intellectual content. K.K., U.U., O.F., M.S., T.C., A.B., S.M.V. and J.M.N. participated in study concept and design, interpretation of data, and critical revision of the manuscript for intellectual content. Å.L., J.T., A.G.Z., J.X.S., W.H., M.R., B.A. and J.K. led the design, funding, and conduct of the TEDDY study, interpreted the data, and critically reviewed the manuscript for intellectual content. All authors approved the final submitted version. The TEDDY Study Group jointly supervised the work.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-51251-4>.

**Competing Interests:** The authors declare no competing interests.

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

## Consortia The TEDDY Study Group

Kimberly Bautista<sup>13</sup>, Judith Baxter<sup>13</sup>, Daniel Felipe-Morales<sup>13</sup>, Kimberly Driscoll<sup>13</sup>, Brigitte I. Frohnert<sup>13</sup>, Marisa Gallant<sup>13</sup>, Patricia Gesualdo<sup>13</sup>, Michelle Hoffman<sup>13</sup>, Rachel Karban<sup>13</sup>, Edwin Liu<sup>13</sup>, Andrea Steck<sup>13</sup>, Kathleen Waugh<sup>13</sup>, Olli G. Simell<sup>9</sup>, Annika Adamsson<sup>19</sup>, Suvi Ahonen<sup>15,16,17</sup>, Mari Åkerlund<sup>15,16,17</sup>, Anne Hekkala<sup>18,20</sup>, Henna Holappa<sup>18,20</sup>, Heikki Hyöty<sup>16,17</sup>, Anni Ikonen<sup>18,20</sup>, Jorma Ilonen<sup>9,21</sup>, Sinikka Jäminki<sup>16,17</sup>, Sanna Jokipuu<sup>19</sup>, Leena Karlsson<sup>19</sup>, Miia Kähönen<sup>18,20</sup>, Mikael Knip<sup>16,17</sup>, Minna-Liisa Koivikko<sup>18,20</sup>, Mirva Koreasalo<sup>15,16,17</sup>, Kalle Kurppa<sup>16,17</sup>, Jarita Kytölä<sup>16,17</sup>, Tiina Latva-aho<sup>18,20</sup>, Katri Lindfors<sup>16</sup>, Maria Lönnrot<sup>16,17</sup>, Elina Mäntymäki<sup>19</sup>, Markus Mattila<sup>16</sup>, Katja Multasuo<sup>18,20</sup>, Teija Mykkänen<sup>18,20</sup>, Tiina Niininen<sup>16,17</sup>, Sari Niinistö<sup>15,17</sup>, Mia Nyblom<sup>16,17</sup>, Sami Oikarinen<sup>16,17</sup>, Paula Ollikainen<sup>18,20</sup>, Sirpa Pohjola<sup>18,20</sup>, Petra Rajala<sup>19</sup>, Jenna Rautanen<sup>15,17</sup>, Anne Riikonen<sup>15,16,17</sup>, Minna Romo<sup>19</sup>, Suvi Ruohonen<sup>19</sup>, Satu Simell<sup>9</sup>, Maija Sjöberg<sup>9,19</sup>, Aino Stenius<sup>18,20</sup>, Päivi Tossavainen<sup>18,20</sup>, Mari Vähä-Mäkilä<sup>19</sup>, Sini Vainionpää<sup>19</sup>, Eeva Varjonen<sup>9,19</sup>, Riitta Veijola<sup>18,20</sup>, Irene Viinikangas<sup>18,20</sup>, Desmond Schatz<sup>22</sup>, Diane Hopkins<sup>11</sup>, Leigh Steed<sup>11</sup>, Jennifer Bryant<sup>11</sup>, Katherine Silvis<sup>11</sup>, Michael Haller<sup>22</sup>, Melissa Gardiner<sup>11</sup>, Richard McIndoe<sup>11</sup>, Ashok Sharma<sup>11</sup>, Stephen W. Anderson<sup>23</sup>, Laura Jacobsen<sup>22</sup>, John Marks<sup>22</sup>, P. D. Towe<sup>22</sup>, Ezio Bonifacio<sup>24</sup>, Miryam D'Angelo<sup>7</sup>, Anita Gavrisan<sup>7</sup>, Cigdem Gezginci<sup>7</sup>, Anja Heublein<sup>7</sup>, Verena Hoffmann<sup>7</sup>, Sandra Hummel<sup>7</sup>, Andrea Keimer<sup>26</sup>, Annette Knopff<sup>7</sup>, Charlotte Koch<sup>7</sup>, Sibylle Koletzko<sup>25</sup>, Claudia Ramminger<sup>7</sup>, Roswith Roth<sup>7</sup>, Marlon Scholz<sup>7</sup>, Joanna Stock<sup>7</sup>, Katharina Warncke<sup>7</sup>, Lorena Wendel<sup>7</sup>, Christiane Winkler<sup>7</sup>, Daniel Agardh<sup>8</sup>, Carin Andréon Aronsson<sup>8</sup>, Maria Ask<sup>8</sup>, Jenny Bremer<sup>8</sup>, Corrado Cilio<sup>8</sup>, Emelie Ericson-Hallström<sup>8</sup>, Annika Fors<sup>8</sup>, Lina Fransson<sup>8</sup>, Thomas Gard<sup>8</sup>, Rasmus Bennet<sup>8</sup>, Monika Hansen<sup>8</sup>, Susanne Hyberg<sup>8</sup>, Hanna Jisser<sup>8</sup>, Fredrik Johansen<sup>8</sup>, Berglind Jonsdottir<sup>8</sup>, Silvija Jovic<sup>8</sup>, Helena Elding Larsson<sup>8</sup>, Marielle Lindström<sup>8</sup>, Markus Lundgren<sup>8</sup>, Maria Månsson-Martinez<sup>8</sup>, Maria Markan<sup>8</sup>, Jessica Melin<sup>8</sup>, Zeliha Mestan<sup>8</sup>, Caroline Nilsson<sup>8</sup>, Karin Ottosson<sup>8</sup>, Kobra Rahmati<sup>8</sup>, Anita Ramelius<sup>8</sup>, Falastin Salami<sup>8</sup>, Anette Sjöberg<sup>8</sup>, Birgitta Sjöberg<sup>8</sup>, Carina Törn<sup>8</sup>, Anne Wallin<sup>8</sup>, Åsa Wimar<sup>8</sup>, Sofie Åberg<sup>8</sup>, Michael Killian<sup>12</sup>, Claire Cowen Crouch<sup>12</sup>, Jennifer Skidmore<sup>12</sup>, Ashley Akramoff<sup>12</sup>, Masumeh Chavoshi<sup>12</sup>, Kayleen Dunson<sup>12</sup>, Rachel Hervey<sup>12</sup>, Rachel Lyons<sup>12</sup>, Arlene Meyer<sup>12</sup>, Denise Mulenga<sup>12</sup>, Jared Radtke<sup>12</sup>, Matei Romancik<sup>12</sup>, Davey Schmitt<sup>12</sup>, Julie Schwabe<sup>12</sup>, Sarah Zink<sup>12</sup>, Sarah Austin-Gonzalez<sup>4</sup>, Maryouri Avendano<sup>4</sup>, Sandra Baethke<sup>4</sup>, Rasheedah Brown<sup>4</sup>, Brant Burkhardt<sup>4</sup>, Martha Butterworth<sup>4</sup>, Joanna Clasen<sup>4</sup>, David Cuthbertson<sup>4</sup>, Christopher Eberhard<sup>4</sup>, Steven Fiske<sup>4</sup>, Jennifer Garmeson<sup>4</sup>, Veena Gowda<sup>4</sup>, Kathleen Heyman<sup>4</sup>, Belinda Hsiao<sup>4</sup>, Christina Karges<sup>4</sup>, Francisco Perez Laras<sup>4</sup>, Hye-Seung Lee<sup>4</sup>, Qian Li<sup>4</sup>, Shu Liu<sup>4</sup>, Xiang Liu<sup>4</sup>, Kristian Lynch<sup>4</sup>, Colleen Maguire<sup>4</sup>, Jamie Malloy<sup>4</sup>, Cristina McCarthy<sup>4</sup>, Aubrie Merrell<sup>4</sup>, Steven Meulemans<sup>4</sup>, Hemang Parikh<sup>4</sup>, Ryan Quigley<sup>4</sup>, Cassandra Remedios<sup>4</sup>, Chris Shaffer<sup>4</sup>, Laura Smith<sup>4</sup>, Susan Smith<sup>4</sup>, Noah Sulman<sup>4</sup>, Roy Tamura<sup>4</sup>, Dena Tewey<sup>4</sup>, Michael Toth<sup>4</sup>, Kendra Vehik<sup>4</sup>, Ponni Vijayakandipan<sup>4</sup>, Keith Wood<sup>4</sup>, Jimin Yang<sup>4</sup>, Liping Yu<sup>13</sup>, Dongmei Miao<sup>13</sup>, Polly Bingley<sup>27</sup>, Alistair Williams<sup>27</sup>, Kyla Chandler<sup>27</sup>, Olivia Ball<sup>27</sup>, Ilana Kelland<sup>27</sup>, Sian Grace<sup>27</sup>, Ben Gillard<sup>27</sup>, Masumeh Chavoshi<sup>12</sup>, Jared Radtke<sup>12</sup>, Julie Schwabe<sup>12</sup>, Bill Wikoff<sup>3</sup>, Dmitry Grapov<sup>3</sup>, Tobias Kind<sup>3</sup>, Mine Palazoglu<sup>3</sup>, Luis Valdiviez<sup>3</sup>, Benjamin Wancewicz<sup>3</sup>, Gert Wohlgemuth<sup>3</sup>, Joyce Wong<sup>3</sup>, Sandra Ke<sup>28</sup>, Niveen Mulholland<sup>28</sup>, Kasia Bourcier<sup>29</sup>, Thomas Briese<sup>30</sup>, Suzanne Bennett Johnson<sup>31</sup> & Eric Triplett<sup>22</sup>

<sup>18</sup>University of Oulu, Oulu, Finland. <sup>19</sup>Turku University Hospital, Hospital District of Southwest Finland, Turku, Finland. <sup>20</sup>Oulu University Hospital, Oulu, Finland. <sup>21</sup>University of Eastern Finland Kuopio, Kuopio, Finland. <sup>22</sup>University of Florida, Gainesville, FL, USA. <sup>23</sup>Pediatric Endocrine Associates, Children's Healthcare of Atlanta, Atlanta, GA, USA. <sup>24</sup>Center for Regenerative Therapies, TU Dresden, Dresden, Germany. <sup>25</sup>Dr. von Hauner Children's Hospital, Department of Gastroenterology, Ludwig Maximilians University Munich, München, Germany. <sup>26</sup>University of Bonn, Department of Nutritional Epidemiology, Bonn, Germany. <sup>27</sup>Bristol Medical School, University of Bristol, Bristol, UK. <sup>28</sup>NIDDK Biosample Repository at Fisher BioServices, Rockville, MD, USA. <sup>29</sup>National Institutes of Allergy and Infectious Diseases, North Bethesda, MD, USA. <sup>30</sup>Columbia University, New York, NY, USA. <sup>31</sup>Florida State University College of Medicine, Tallahassee, FL, USA.