

# SCIENTIFIC DATA

OPEN

## Data Descriptor: A longitudinal plasma lipidomics dataset from children who developed islet autoimmunity and type 1 diabetes

Received: 1 June 2018  
Accepted: 26 September 2018  
Published: 13 November 2018

Santosh Lamichhane<sup>1</sup>, Linda Ahonen<sup>2</sup>, Thomas Sparholt Dyrland<sup>2</sup>, Heli Siljander<sup>3,4</sup>, Heikki Hyöty<sup>5,6</sup>, Jorma Ilonen<sup>7,8</sup>, Jorma Toppari<sup>9,10</sup>, Riitta Veijola<sup>11,12,13</sup>, Tuulia Hyötyläinen<sup>14</sup>, Mikael Knip<sup>3,4,15,16</sup> & Matej Orešič<sup>1,17</sup>

Early prediction and prevention of type 1 diabetes (T1D) are currently unmet medical needs. Previous metabolomics studies suggest that children who develop T1D are characterised by a distinct metabolic profile already detectable during infancy, prior to the onset of islet autoimmunity. However, the specificity of persistent metabolic disturbances in relation T1D development has not yet been established. Here, we report a longitudinal plasma lipidomics dataset from (1) 40 children who progressed to T1D during follow-up, (2) 40 children who developed single islet autoantibody but did not develop T1D and (3) 40 matched controls (6 time points: 3, 6, 12, 18, 24 and 36 months of age). This dataset may help other researchers in studying age-dependent progression of islet autoimmunity and T1D as well as of the age-dependence of lipidomic profiles in general. Alternatively, this dataset could more broadly used for the development of methods for the analysis of longitudinal multivariate data.

<b>Design Type(s)</b>	disease state design • observational design • longitudinal study design
<b>Measurement Type(s)</b>	lipid
<b>Technology Type(s)</b>	liquid chromatography-tandem mass spectrometry
<b>Factor Type(s)</b>	age • disease stage • autoantibody measurement
<b>Sample Characteristic(s)</b>	Homo sapiens • blood plasma

<sup>1</sup>Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku 20520, Finland. <sup>2</sup>Steno Diabetes Center Copenhagen, 2820 Gentofte, Denmark. <sup>3</sup>Children's Hospital, University of Helsinki and Helsinki University Hospital, 00290 Helsinki, Finland. <sup>4</sup>Research Program Unit, Diabetes and Obesity, University of Helsinki, 00290 Helsinki, Finland. <sup>5</sup>Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. <sup>6</sup>Fimlab Laboratories, Pirkanmaa Hospital District, Tampere, Finland. <sup>7</sup>Immunogenetics Laboratory, Institute of Biomedicine, University of Turku, Turku, Finland. <sup>8</sup>Clinical Microbiology, Turku University Hospital, Turku, Finland. <sup>9</sup>Institute of Biomedicine, Centre for Integrative Physiology and Pharmacology, University of Turku, Turku, Finland. <sup>10</sup>Department of Pediatrics, Turku University Hospital, Turku, Finland. <sup>11</sup>Department of Paediatrics, PEDEGO Research Unit, Medical Research Centre, University of Oulu, Oulu, Finland. <sup>12</sup>Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland. <sup>13</sup>Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden. <sup>14</sup>Department of Chemistry, Örebro University, 702 81 Örebro, Sweden. <sup>15</sup>Tampere Center for Child Health Research, Tampere University Hospital, Tampere, Finland. <sup>16</sup>Folkhälsan Research Center, Helsinki, Finland. <sup>17</sup>School of Medical Sciences, Örebro University, 702 81 Örebro, Sweden. Correspondence and requests for materials should be addressed to M.K. (email: mikael.knip@helsinki.fi) or M.O. (email: matej.oresic@utu.fi)

## Background & Summary

Type 1 diabetes (T1D) is a chronic, autoimmune disease caused by progressive loss of insulin-secreting capacity due to the selective death of beta cells in the islets of Langerhans, of which there are more than one million in the human pancreas<sup>1</sup>. The age of onset is usually between 5–15 years, but, in recent years, many children before 5 years of age are being affected<sup>2</sup>. Although approximately 80% of subjects with T1D carry defined risk-associated genotypes at the human leukocyte antigen (HLA) locus, only 3–7% of the carriers of such genetic risk markers go on to develop overt disease. Seroconversion to islet autoantibody positivity is the first detectable signal demonstrating initiation of autoimmunity and risk of progression towards diabetes<sup>3,4</sup>. However, whilst seroconversion to autoantibody positivity precedes clinical disease by months to years, the point at which seroconversion occurs may already be too late for therapeutic approaches aimed at preventing progression to overt diabetes.

Previous metabolomic studies suggest that children who develop islet autoimmunity and/or T1D are characterised by specific metabolic disturbances prior to the first appearance of islet autoantibodies<sup>5–7</sup>. Metabolic profiling may thus be of clinical relevance by providing a complimentary tool for estimating risk of progression to T1D. However, the underlying causes of these early metabolic disturbances and their link to disease progression are still largely unknown. It is also not yet established if the observed, persistent metabolic changes are specifically associated with the progression to T1D, or more broadly, if they are associated with progression to islet autoimmunity and irrespective of disease outcome.

Here we provide a longitudinal plasma lipidomics dataset obtained from children participating in a prospective, birth cohort Type 1 Diabetes Prediction and Prevention study (DIPP). Three study groups were examined: children who (1) progressed to T1D (PT1D), (2) developed at least a single islet autoantibody (Ab) during follow-up but did not progress to T1D (P1Ab), and (3) controls (CTR) who remained autoantibody negative and healthy during the follow-up until 15 years of age. We analysed 428 plasma samples from 120 children (40 PT1D, 40 P1Ab and 40 CTR). The samples were collected up to six different time points corresponding to the ages of 3, 6, 12, 18, 24, and 36 (or above) months (Fig. 1). These age groups were selected with the objective of understanding the changes in lipidomic profile preceding prior to overt T1D. We performed untargeted lipidomics using ultra-high-performance liquid chromatography combined with quadrupole time-of-flight mass spectrometry (UHPLC-QTOF-MS). Both raw and pre-processed datasets were deposited in the MetaboLights repository (Data Citation 1). Along with the aforementioned lipidomic data, we provide information on the type of islet autoantibodies observed in this longitudinal setting. In this study, sphingomyelins (SMs) were found to be persistently downregulated in PT1D as compared to the P1Ab and CTR groups<sup>8</sup>. Triacylglycerols (TGs) and phosphatidylcholines (PCs) were mainly downregulated in PT1D as compared to P1Ab at the age of 3 months. These results suggest that distinct lipidomic signatures characterise children who progressed to islet autoimmunity or clinical T1D. Lipidomic profiling may thus be helpful in the identification of at-risk children before the initiation of autoimmunity.

This data descriptor is one of the first longitudinal lipidomic datasets allowing the investigation of progression to islet autoimmunity/T1D during the early prodromal phases of disease development. Considering the longitudinal study design, this clinical dataset may have many uses. Firstly, it may assist other researchers in their studies of early pathogenesis of T1D and, potentially, other immune-mediated inflammatory diseases. It may also allow other researchers to study the age-dependent progression of lipidomic profiles during infancy. Finally, the dataset has great potential to be used in the development and testing of algorithms for the analysis of multivariate longitudinal/prospective data.

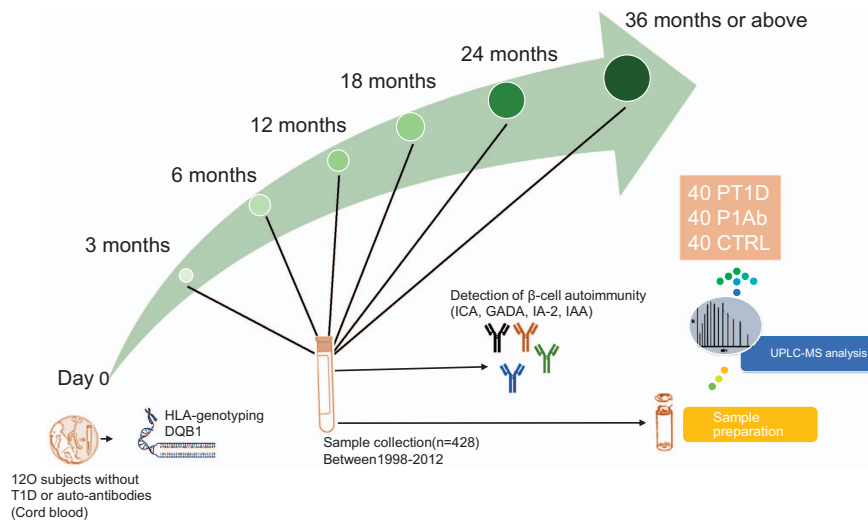
## Methods

These methods are expanded versions of descriptions in our related work<sup>8</sup>.

### Study design

The plasma samples for lipidomics analysis were obtained from the Finnish Type 1 Diabetes Prevention and Prediction (DIPP) study<sup>9</sup>. The DIPP study has screened more than 220,000 newborn infants for HLA-conferred susceptibility to T1D at three university hospitals in Finland: Turku, Tampere and Oulu<sup>10</sup>. Over 25,000 infants were identified as having an increased genetic risk and approximately 17,000 families joined the follow-up study, which involves regular study center visits (with an interval of 3–6 months). The Ethics and Research Committee of the participating Universities and Hospitals approved the study protocol. The study was conducted according to the guidelines in the Declaration of Helsinki. All families provided written informed consent for participation in the study. At every visit, blood samples were collected to measure the titre of T1D-associated islet autoantibodies. Non-fasting blood samples were collected into sodium citrate tubes. Plasma was separated within 30 min of collection by centrifugation at 1600g for 20 min at room temperature, aliquoted, and stored at –80 °C until analysed. Now, 1,663 of those children (9.8%) have seroconverted to positivity for one autoantibody during the follow-up, 808 (4.8%) have developed multiple autoantibodies, and 510 (3.1%) have progressed to clinical T1D.

The 120 infants included in the current data descriptor were selected from a subset of DIPP children from the city of Tampere. Up to six longitudinal samples per child were collected between 1998 and 2012, corresponding to the ages of 3, 6, 12, 18, 24 and 36 months (Fig. 1, Table 1). The details of the ages and selected characteristics of the current study subjects are given in the metadata (Data Citation 1). The three



**Figure 1.** An overview of the study design. In this study, newborn infants at the University Hospital of Tampere were screened for increased genetic risk for T1D. This study includes three study groups: (1) 40 T1D (PT1D) progressors, (2) 40 who developed at least a single Ab during the follow-up but did not progress to T1D (P1Ab), and (3) 40 controls (CTRL) who remained autoantibody negative during the follow-up. The three study groups were matched by HLA-associated diabetes risk, gender and period of birth. For each child, up to six longitudinal samples were obtained corresponding to the ages of 3, 6, 12, 18, 24 and 36 (or above) months. Abbreviations: islet cell antibodies (ICA), insulin autoantibodies (IAA), islet antigen 2 (IA-2) autoantibodies, and glutamic acid decarboxylase autoantibodies (GADA).

	PT1D (n = 40)	P1Ab (n = 40)	CTR (n = 40)
Gender (girls, boys)	(14, 26)	(14, 26)	(14, 26)
Age at time of diagnosis (years) (mean $\pm$ SD)	4.75 $\pm$ 2.94	NA	NA
Age at time of first seroconversion (years) (mean $\pm$ SD)	1.34 $\pm$ 0.58	3.05 $\pm$ 2.50	NA

**Table 1.** Anthropometric characteristics of study population.

study groups were matched by HLA-associated diabetes risk, gender and the period of birth. In total, 428 plasma samples were selected and analysed for this study.

### HLA genotyping

HLA-conferred susceptibility to T1D was analysed using cord blood samples as previously described<sup>11</sup>. Briefly, the HLA-genotyping was performed with a time-resolved fluorometry-based assay for four alleles using lanthanide chelate labeled, sequence-specific oligonucleotide probes detecting the DQB1\*02, DQB1\*03:01, DQB1\*03:02 and DQB1\*06:02/3 alleles<sup>12</sup>. The carriers of the genotype DQB1\*02/DQB1\*03:02 or DQB1\*03:02/x genotypes (here  $x \neq$  DQB1\*02, DQB1\*03:01, DQB1\*06:02, or DQB1\*06:03 alleles) were categorised into the T1D risk group and recruited for the DIPP follow-up program.

### Detection of beta-cell autoimmunity

The participants with HLA-conferred genetic susceptibility were monitored for the appearance of T1D-associated autoantibodies: islet cell antibodies (ICA), insulin autoantibodies (IAA), islet antigen 2 autoantibodies (IA-2A), and glutamic acid decarboxylase autoantibodies (GADA). These autoantibodies were measured in the Diabetes Research Laboratory (University of Oulu) from the plasma samples taken at each follow-up visit<sup>13</sup>. ICA were detected with the use of indirect immunofluorescence, whereas the other three autoantibodies were quantified with the use of specific radiobinding assays<sup>14</sup>. We used cutoff limits for positivity of 2.5 Juvenile Diabetes Foundation (JDF) units for ICA, 3.48 relative units (RU) for IAA, 5.36 RU for GADA, and 0.43 RU for IA-2A. The disease sensitivity and specificity of the assay for ICA were 100% and 98%, respectively, in the fourth round of the international workshops on standardisation of the ICA assay. According to the Diabetes Autoantibody standardisation Program

(DASP) and Islet Autoantibody standardisation Program (IASP) workshop results in 2010–2015, disease sensitivities for the IAA, GADA and IA-2A radio binding assays were 36–62%, 64–88% and 62–72%, respectively. The corresponding disease specificities were 94–98%, 94–99% and 93–100%, respectively. The metadata in the Metabolights contains information on the types of autoantibodies detected in the longitudinal plasma samples (Data Citation 1).

### Sample preparation and UHPLC-MS analysis

The solvents and eluent modifiers used in this work were purchased from Sigma-Aldrich (Steinheim, Germany): HPLC-grade chloroform, LC-MS grade acetonitrile (ACN), isopropanol (IPA), water (H<sub>2</sub>O), methanol (MeOH), ammonium acetate (NH<sub>4</sub>Ac), analytical grade formic acid (HCOOH) and sodium chloride (NaCl).

The following internal standards were purchased for quality control (QC) and calibration purposes: 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)) from Avanti Polar Lipids, Inc. (Alabaster, Alabama, USA), 1,2-Dimyristoyl-*sn*-glycero-3-phospho(choline-d<sub>13</sub>) (PC(14:0/d13)) from Sigma-Aldrich and Tripalmitin-1,1,1-13C<sub>3</sub> (TG(16:0/16:0/16:0)-13C<sub>3</sub>) and Trioctanoin-1,1,1-13C<sub>3</sub> (TG(8:0/8:0/8:0)-13C<sub>3</sub>) from Larodan (Solna, Sweden). Stock solutions (1.0 mg mL<sup>-1</sup>) of the internal standards were prepared by dissolving the analytes in CHCl<sub>3</sub>:MeOH (2:1, v/v). The working standard solutions were prepared by further diluting the stock solutions in CHCl<sub>3</sub>:MeOH (2:1, v/v) to achieve concentrations of 250 ng mL<sup>-1</sup> and 3.5 µg mL<sup>-1</sup>.

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-octadecanoyl-*sn*-glycero-3-phosphocholine (LPC(18:0)), 1-(1Z-octadecenyl)-2-docosahexaenoyl-*sn*-glycero-3-phosphocholine (PC(18:0p/22:6)) and 1-stearoyl-2-linoleoyl-*sn*-glycerol (DG(18:0/20:4)) from Avanti Polar Lipids Inc., 1-Palmitoyl-2-Hydroxy-*sn*-Glycero-3-Phosphatidylcholine (LPC(16:0)) from Larodan, and 1,2,3-Triheptadecanoylglycerol (TG(17:0/17:0/17:0)) and 3β-Hydroxy-5-cholestene 3-linoleate (ChoE(18:2)) from Sigma-Aldrich were prepared to the following concentration levels: 100, 500, 1000, 1500, 2000 and 2500 ng mL<sup>-1</sup> (in CHCl<sub>3</sub>:MeOH, 2:1, v/v) including 250 ng mL<sup>-1</sup> of each QC standard.

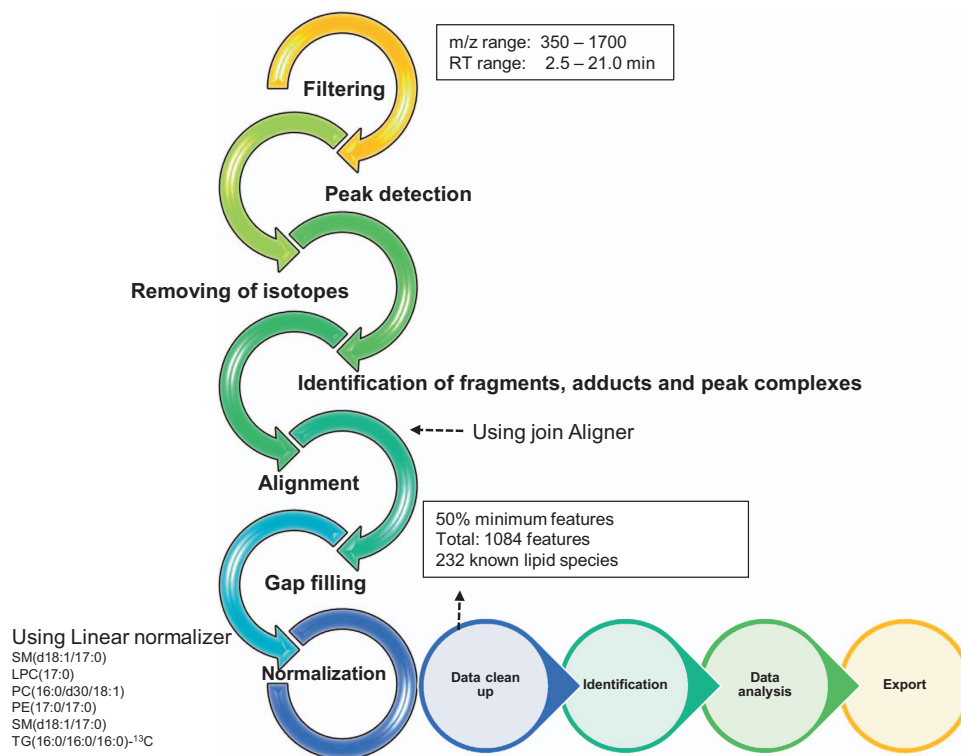
A total of 428 plasma samples were extracted in randomised order using a modified version of the Folch procedure: 10 µL of 0.9% NaCl, 40 µL of CHCl<sub>3</sub>:MeOH (2:1, v/v) and 80 µL of the 3.5 µg mL<sup>-1</sup> working standards solution were added to 10 µL of each plasma sample. The samples were vortex mixed and incubated on ice for 30 min after which they were centrifuged (9400 × g, 3 min, 4 °C). From the lower layer of each sample, 60 µL was transferred to a glass vial with an insert and 60 µL of CHCl<sub>3</sub>:MeOH (2:1, v/v) and added to each sample. The samples were re-randomised and stored at -80 °C until analysis on the UHPLC-QTOF-MS system.

The UHPLC system used was a 1290 Infinity system from Agilent Technologies (Santa Clara, California, USA). The system was equipped with a multisampler (maintained at 10 °C) using 10% DCM in MeOH and ACN:MeOH:IPA:H<sub>2</sub>O (1:1:1:1, v/v/v/v) + 0.1% HCOOH as needle wash solutions after each injection for 7.5 s each, a quaternary solvent manager and a column thermostat (maintained at 50 °C). Separations were performed on an ACQUITY UPLC<sup>®</sup> BEH C18 column (2.1 mm × 100 mm, particle size 1.7 µm) by Waters (Milford, USA). The flow rate was 0.4 mL min<sup>-1</sup> and the injection volume was 1 µL. H<sub>2</sub>O + 1% NH<sub>4</sub>Ac (1 M) + 0.1% HCOOH (A) and ACN:IPA (1:1, v/v) + 1% NH<sub>4</sub>Ac + 0.1% HCOOH (B) were used as the mobile phases for the gradient elution. The gradient was as follows: from 0 to 2 min. 35–80% B, from 2 to 7 min. 80–100% B and from 7 to 14 min 100% B. Each run was followed by a 7 min re-equilibration period under the initial conditions (35% B).

The mass spectrometer coupled to the UHPLC was a 6550 iFunnel QTOF-MS from Agilent Technologies interfaced with a dual jet stream electrospray (dual ESI) ion source. Nitrogen generated by a nitrogen generator (PEAK Scientific, Scotland, UK) was used as the nebulising gas at a pressure of 21 psi, as the drying gas at a flow rate of 14 L min<sup>-1</sup> (at 193 °C) and as the sheath gas at a flow rate of 11 L min<sup>-1</sup> (at 379 °C). Pure nitrogen (6.0) from Praxair (Fredericia, Denmark) was used as the collision gas. The capillary voltage and the nozzle voltage were kept at 3643 V and 1500 V, respectively. The reference mass solution including ions at *m/z* 121.0509 and 922.0098 was prepared according to instructions by Agilent and was introduced to the mass spectrometer through the other nebuliser in the dual ESI ion source using a separate Agilent series 1290 isocratic pump at a constant flow rate of 4 mL min<sup>-1</sup> (split to 1:100 before the nebuliser). The acquisition mass range was *m/z* 100–1700 and the instrument was run in extended dynamic range mode with an approximate resolution of 30,000 FWHM measured at *m/z* 1521.9715 (which is included in the tune mixture) during calibration of the instrument. MassHunter B.06.01 software (Agilent) was used for data acquisition.

### Data pre-processing

The data pre-processing was performed using MZmine 2.18.2<sup>15</sup>. Here, we have adhered to the data processing steps as suggested by the metabolomics standards initiative<sup>16</sup>. The typical pre-processing workflow includes raw file import, filtering/smoothing, detection of peaks, peak list de-isotoping,



**Figure 2.** The typical workflow for processing mass spectrometry data using MZmine 2. The workflow includes raw file import, filtering/smoothing, detection of peaks, peak list de-isotoping, alignment, gap filling, and integration of peaks, normalisation, and, finally, peak/feature identification.

alignment, gap filling, integration of peaks, normalisation and, finally, peak/feature identification (Fig. 2, and Supplementary Table S1). The following steps were applied in the processing: (1) crop filtering with a  $m/z$  range of 350–1700  $m/z$  and a retention time (RT) range of 2.5 to 21.0 min, (2) mass detection with a noise level of 750, (3) chromatogram builder with a minimum time span of 0.08 min, minimum height of 2250 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, 2250 minimum absolute height, a minimum ration of peak top/edge of 1 and a peak duration range of 0.08 to 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) peak list row filter keeping only peak with a minimum of 1 peak in a row, (8) join aligner with a  $m/z$  tolerance of 0.006 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, (9) peak list row filter with a minimum of 53 peak in a row (=10% of the samples), (10) duplicate peak filter with a  $m/z$  tolerance of 0.006  $m/z$  or 10.0 ppm and a RT tolerance of 0.1 min, (11) gap filling using the same RT and  $m/z$  range gap filler algorithm with an  $m/z$  tolerance of 0.006  $m/z$  or 10.0 ppm, (12) peak filter using the same parameters as under step 6, (13) peak list row filter with a minimum of 265 peaks in a row (=50% of the samples), enabling detection of 1084 features (14) identification of lipids using a custom database search (232 identified out of 1084 features) with an  $m/z$  tolerance of 0.006  $m/z$  or 10.0 ppm and a RT tolerance of 0.1 min, (15) normalisation using lipid-class-specific internal standards (PE (17:0/17:0), SM (d18:1/17:0), Cer (d18:1/17:0), LPC (17:0), TG (16:0/16:0/16:0)-<sup>13</sup>C3 and PC (16:0/d30/18:1)) using in-house developed R-script, and (16) data imputation of missing values with the half of the minimum value for each lipid.

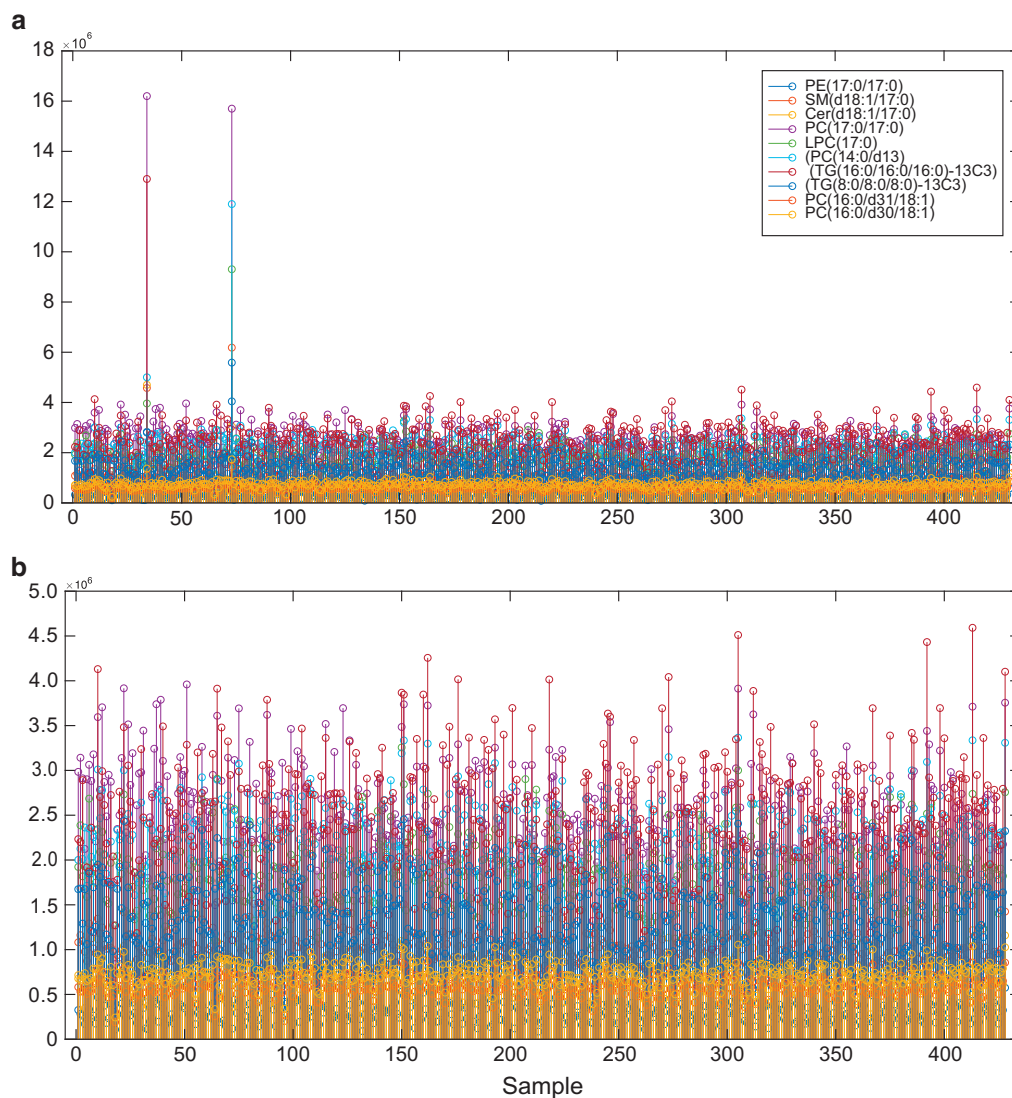
#### Code availability

All the pre-processing analyses were performed using the publicly-available software package MZmine 2.18.2 and with parameters as described in Supplementary Table S1.

#### Data records

The raw data files in .mzML format are deposited in the MetaboLights repository (Data Citation 1). Additionally, the deposited data contains the pre-processed lipidomic data as obtained from MZmine





**Figure 3.** Quality control of UHPLC-QTOFMS run evaluated using peak areas from nine internal standard compounds. (a) Plot of peak areas with 430 samples. (b) Plot of peak areas of 428 samples, *i.e.* after removal of 2 outlier samples. Here, each color represents the peak area of the given internal standard PE (17:0/17:0), SM(d18:1/17:0), Cer (d18:1/17:0), PC(17:0/17:0), LPC(17:0), PC(14:0/d13), TG(16:0/16:0/16:0)-13C3, TG(8:0/8:0/8:0)-13C3 and PC(16:0/d31/18:1).

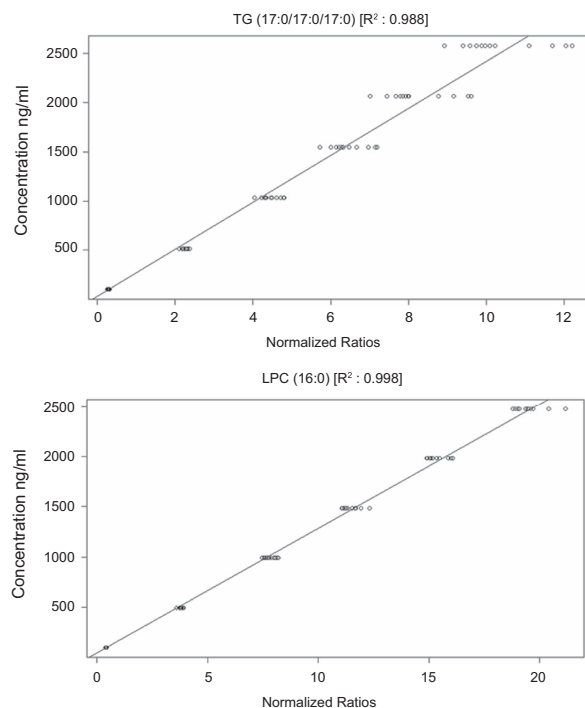
data processing, including the identified lipids. The associated data were captured using the ISA-creator package available from MetaboLights. This manuscript describes the samples, data collection, processing steps and overall study design.

#### Ethical approval and informed consent

The ethics and research committees of both the participating university and hospital at University of Tampere, Tampere Finland, approved the study protocol. The study was conducted according to the guidelines in the Declaration of Helsinki. All families provided written informed consent for participation in the study.

#### Technical validation

This study followed the best practices of analytical methodologies for the global profiling of lipids in the plasma sample as described by Hyotylainen *et al.*<sup>17,18</sup>. For quality monitoring, QC samples (pooled plasma samples), blank samples, pure standards (standards in solvent) were run at regular intervals after every 4–16 study samples. To assure data quality, relative standard deviations (%RSDs) were calculated for the retention times and the peak areas for each QC standard in the QC samples



**Figure 4. Representative global linear models.** Six different concentrations of lipid class-specific standards TG (17:0/17:0/17:0), and LPC (16:0) were used to generate the calibration plot.

throughout the sample set. Overall, these results show a reproducible method. For example, pooled plasma samples analysed throughout the sample set show average %RSDs for the four batches of samples of less than 0.2 and 15.4% (normalised using PC(16:0/d30/18:1)) for the retention times and peak areas, respectively. The %RSDs for the rest of the QC samples were in the same range as for the pooled plasma samples.

To assess experimental error, nine internal standards were spiked into the plasma samples during sample preparation. The plot of peak areas from nine QC standards (Fig. 3a) clearly shows that two samples had eccentric internal standard profiles. This was traced to an error during sample preparation, wherein the QC standard was added two these two samples twice. Therefore, these two outlier samples were not included in downstream analysis (and not included in the data descriptor). The representative quality plots after removal of aberrant samples (Fig. 3b) demonstrates that the peak areas of the QC standards have very low variability between the measured samples. This indicates that use of proper internal standards and careful follow-up of the response of these internal standards critically reduces the chances of unexpected variation, which may arise in LC-MS analysis. Based on this, we conclude that the samples in this data descriptor were successfully analysed following generally accepted analytical quality guidelines.

During LC-MS analyses, it is also highly important to keep track of blank samples inserted into the sample sequence. This is done to detect possible contamination/interference, which affect the actual sample data. We analysed 293 blank or solvent blank samples throughout this sample set. Solvent blank samples (consisting of a mixture of CHCl<sub>3</sub>: MeOH, 2:1, v/v) were analysed after every 4<sup>th</sup> sample and blank samples (the same solvent mixture, which underwent the normal sample preparation procedure except for the addition of internal standards) after every 16<sup>th</sup> sample. Additionally, a set of QC samples (186 samples) were analysed at the same frequency as the blank samples (*i.e.* after every 16<sup>th</sup> sample). This set of QC samples consisted of: (1) extracted and non-extracted standard samples (the chosen QC standards in solvent), (2) a pooled plasma sample originating from a larger plasma pool from actual diabetes patients for following up on long-term method performance and to facilitate batch correction if needed, and (3) a Standard Reference Material 1950 sample from the National Institute of Standards and Technology (NIST) that serve as community-wide benchmarks for intra- and inter-laboratory QC and method validation<sup>19</sup>.

Next, for quantitative analysis, calibration curves of each analyte must be generated. For global lipidomics analyses this is, however, logistically and practically impossible due to the high number of analytes present in the biological sample matrix. To overcome this limitation, lipid-class-specific compounds were used for creating a semi-quantitative method including six different concentrations of the lipid class specific standards PC (16:0e/18:1(9Z)), TG (17:0/17:0/17:0), LPC (18:0), DG (18:0/20:4) and CE (18:2). The acceptance of LC-MS data usually relies on QC samples being

successfully quantified within pre-defined limits. Achievement of this standard is evaluated statistically (global linear models) based on the value for the coefficient of determination ( $R^2$ ). Figure 4 shows coefficients of determination higher than 0.9 for TG (17:0/17:0/17:0) and LPC(16:0) at six different concentrations using a global linear model with a 1/x weighting. The other standards show similar values (i.e.  $R^2 > 0.9$ ). This statistical analysis consequently ensured the reliability of results presented in this data descriptor.

Furthermore, the assurances provided by the QC methods herein derive from analytical reproducibility, within or between batches of the LC-MS data. These inter- and intra-batch variations can be assessed and corrected using the QC samples. The principal component analysis (PCA) score plot (Supplementary Figure S1) highlights the aforementioned reproducibility across all batches obtained from the pre-processed dataset in this data descriptor. From this score plot, it is apparent that there are no patterns or any clustering that correlated with the batch run during LC-MS. Together, the dataset in this data descriptor reflects robustness of lipid measurement over time, which, in turn, indicates that data analysis performed using this dataset will not be affected by systematic error arising due to instrumental factors.

### Usage notes

The pre-processed data are available from MetaboLights (Data Citation 1). The detail parameters for data pre-processing are available in supplementary material (Supplementary Table S1). The pre-processed data were normalised using the nine different class-specific internal standards. The class-specific were based on PE (17:0/17:0), SM (d18:1/17:0), Cer (d18:1/17:0), LPC (17:0), TG (16:0/16:0/16:0)-13C3 and PC (16:0/d30/18:1), while others were based on retention time; LPC (17:0) for  $\leq 6.0$  min elutions, PC(16:0/d30/18:1) for  $>6.0$ ,  $< 9.0$  min elutions and TG(16:0/16:0/16:0)-13C3 for  $\geq 9.0$  min elutions. In addition, the zero values in the lipidomics dataset were imputed with half of that row's minimum value for each lipid (the code for this imputation were written in-house using R, and are available on request). Due to the non-normal distribution of data, it is recommended that the data are log-transformed before analysis (log-transformation performed in MZmine 2.18.2). Other required guidelines for data-preprocessing are available from the MZmine website (<http://mzmine.github.io/documentation.html>) and authors can be contacted for further help if required.

### References

- Achenbach, P., Bonifacio, E., Koczwara, K. & Ziegler, A.-G. Natural history of type 1 diabetes. *Diabetes* **54**, S25–31 (2005).
- Patterson, C. C., Dahlquist, G. G., 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).
- Kimpimäki, T. *et al.* Natural history of beta-cell autoimmunity in young children with increased genetic susceptibility to type 1 diabetes recruited from the general population. *J. Clin. Endocrinol. Metab.* **87**, 4572–4579 (2002).
- Ziegler, A. G. *et al.* Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA* **309**, 2473–2479 (2013).
- La Torre, D. *et al.* Decreased Cord-Blood Phospholipids in Young Age-at-Onset Type 1 Diabetes. *Diabetes* **62**, 3951–3956 (2013).
- Oresic, M. *et al.* Cord serum lipidome in prediction of islet autoimmunity and type 1 diabetes. *Diabetes* **62**, 3268–3274 (2013).
- Oresic, 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).
- 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).
- Kupila, A. *et al.* Feasibility of genetic and immunological prediction of type I diabetes in a population-based birth cohort. *Diabetologia* **44**, 290–297 (2001).
- Haller, M. J. & Schatz, D. A. The DIPP project: 20 years of discovery in type 1 diabetes. *Pediatr. Diabetes* **17**(Suppl 22): 5–7 (2016).
- Nejentsev, S. *et al.* Population-based genetic screening for the estimation of Type 1 diabetes mellitus risk in Finland: selective genotyping of markers in the HLA-DQB1, HLA-DQA1 and HLA-DRB1 loci. *Diabet. Med.* **16**, 985–992 (1999).
- Ilonen, J. *et al.* Rapid HLA-DQB1 genotyping for four alleles in the assessment of risk for IDDM in the Finnish population. The Childhood Diabetes in Finland (DiMe) Study Group. *Diabetes Care* **19**, 795–800 (1996).
- Siljander, H. T. *et al.* Predictive characteristics of diabetes-associated autoantibodies among children with HLA-conferred disease susceptibility in the general population. *Diabetes* **58**, 2835–2842 (2009).
- Knip, M. *et al.* Dietary intervention in infancy and later signs of beta-cell autoimmunity. *N. Engl. J. Med.* **363**, 1900–1908 (2010).
- 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).
- Goodacre, R. *et al.* Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **3**, 231–241 (2007).
- Hyötyläinen, T., Ahonen, L., Poho, P. & Oresic, M. Lipidomics in biomedical research-practical considerations. *Biochim. Biophys. Acta* **1862**, 800–803 (2017).
- Hyötyläinen, T. & Oresic, M. Bioanalytical techniques in nontargeted clinical lipidomics. *Bioanalysis* **8**, 351–364 (2016).
- Bowden, J. A. *et al.* Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-Metabolites in Frozen Human Plasma. *J. Lipid Res.* **58**, 2275–2288 (2017).

### Data Citations

- Lamichhane, S., Knip, M. & Oresic, M. *MetaboLights* MTBLS620 (2018).

### Acknowledgements

This work was supported by the JDRF grants 4-1998-274, 4-1999-731 4-2001-435 and special research funds for Oulu, Tampere and Turku University Hospitals in Finland. This work was supported by the



Juvenile Diabetes Research Foundation (2-SRA-2014-159-Q-R to M.O.) and the Academy of Finland (Centre of Excellence in Molecular Systems Immunology and Physiology Research – SyMMyS, Decision No. 250114, to M.O. and M.K.). We give our thanks to Olli Simell for his contribution to the DIPP study, Anette Untermann for excellent technical support in metabolomics analysis and Aidan McGlinchey for his contribution in proofreading the manuscript.

### Author Contributions

M.O. and M.K. designed and supervised the study. L.A. and T.S.D. performed metabolomics analysis, which was supervised by T.H.; H.H., J.I., J.T. and R.V. contributed to the design of the clinical study. S.L. and M.O. wrote the manuscript. All authors critically reviewed and approved the final manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/sdata>

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

**How to cite this article:** Lamichhane, S. *et al.* A longitudinal plasma lipidomics dataset from children who developed islet autoimmunity and type 1 diabetes. *Sci. Data.* 5:180250 doi: 10.1038/sdata.2018.250 (2018).

**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 Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.

© The Author(s) 2018