

Persistent alterations in plasma lipid profiles prior to introduction of gluten in the diet associate with progression to celiac disease

Partho Sen, Cecilia Carlsson, Suvi M. Virtanen, Satu Simell, Heikki Hyöty, Jorma Ilonen, Jorma Toppari, Riitta Veijola, Tuulia Hyötyläinen, Mikael Knip, Matej Orešič

Supplementary Notes

Clinical characteristics of the CD progressors and healthy controls

Selected clinical variables measured in the healthy children and biopsy proven CD progressors during the follow-up period are shown in **Figure S1**.

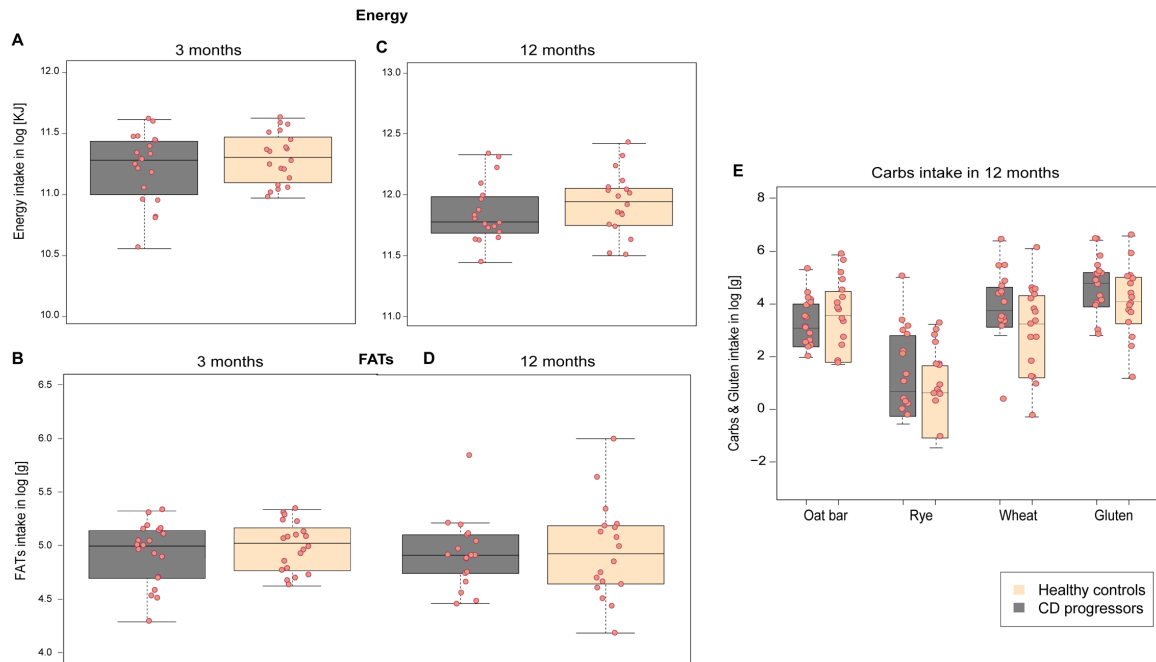


Figure S1 Boxplot showing clinical variables such as energy intake levels (KJ), total carbohydrate (carbs, g), gluten (g) and fatty acids intake levels (g) in CD progressors (grey) and their matched healthy controls (orange) at 3 and/or 12 months. No significant differences were marked (ANOVA, p -values > 0.05) between any of these paired variables.

Change in the fatty acid levels of the CD progressors and healthy controls

Fold changes between CD progressors and healthy controls (CTRL) at the individual fatty acid (FA) level at the age of 3 and 12 months is shown in **Figure S2**.

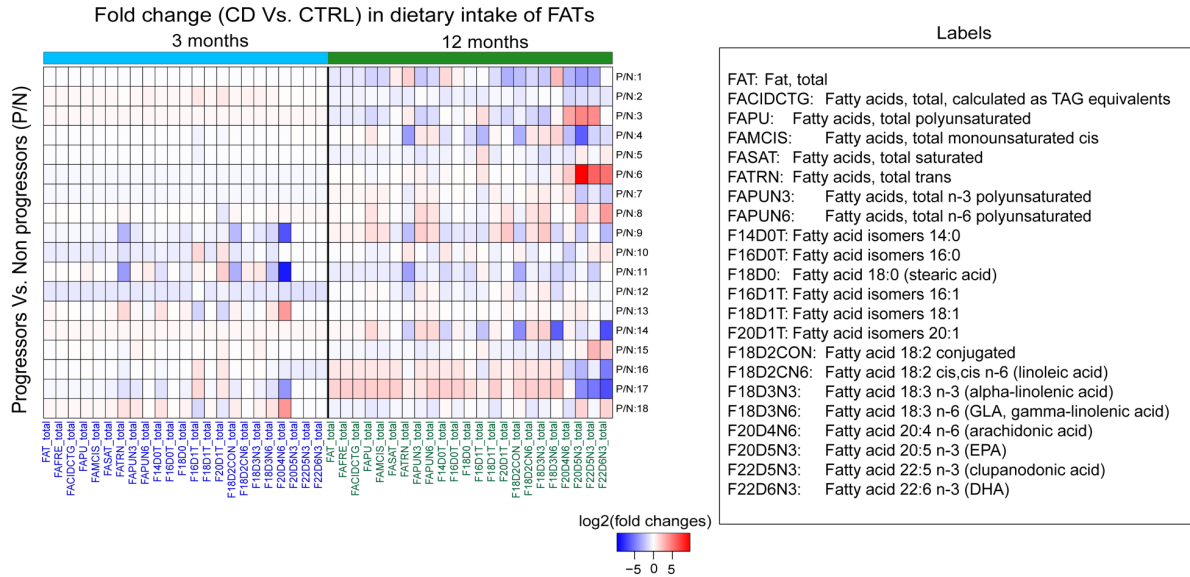


Figure S2 Heatmap showing log₂ fold changes (P/N) in individual fatty acid level (mg) of CD progressor (P) and non-progressor (N) at 3 and 12 months of age. Red color denotes up-regulated in the progressor w.r.t non-progressors, blue denotes down-regulated and white depicts no change or undetected.

Clinical characteristics of the mothers

Clinical variables measured or estimated in the mothers of the CD progressors and healthy controls during pregnancy and lactation are shown in **Figure S3**.

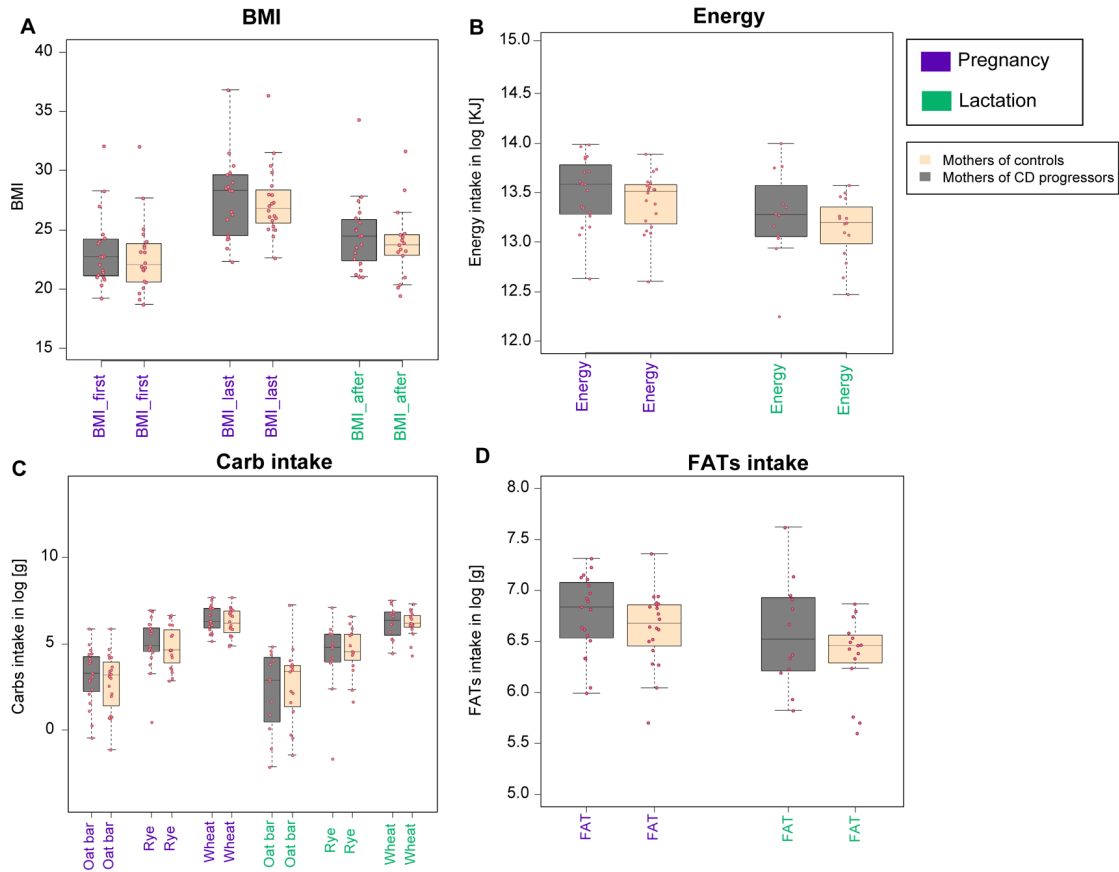


Figure S3 Boxplot showing clinical variables such as body mass index (BMI), total energy intake levels (KJ), carbohydrate (carbs, g) and fatty acids intake levels (g) of mothers of CD progressors (grey) and healthy controls (orange) during pregnancy (blue) and lactation (green). No significant differences observed (ANOVA, $p > 0.05$) between these paired variables.

Similarly, change in the individual FA level was measured in the mothers of CD progressors and healthy controls (CTRL) during pregnancy and lactation (**Figure S4**).

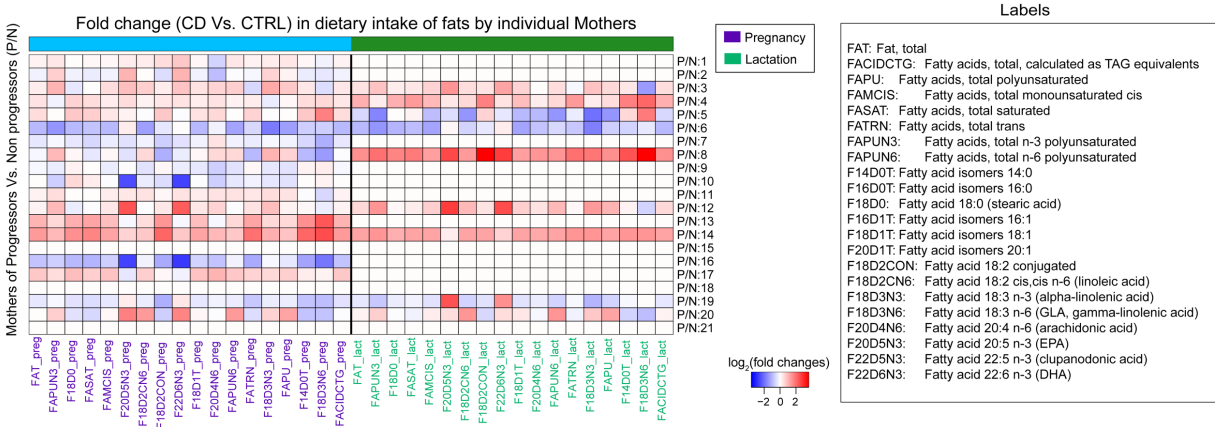


Figure S4 Heatmap showing log₂ fold changes (P/N) in individual fatty acid (mg) intake of the mothers of CD progressor (P) and non-progressor (N)/healthy control (CTRL) during pregnancy (blue) and lactation (green). Red color denotes up-regulated in the progressor, blue denotes down-regulated and white depicts no change or undetected.

Analysis of molecular lipids

A total of 233 plasma samples were randomized and extracted using a modified version of the Folch procedure(1). Promptly after extraction, 10 μL of 0.9% NaCl and 120 μL of CHCl_3 : MeOH (2:1, v/v) containing 2.5 $\mu\text{g mL}^{-1}$ internal standard solution (for quality control (QC) and normalization purposes) were added to 10 μL of each plasma sample. The standard solution contained the following compounds: 1,2-diheptadecanoyl-*sn*-glycero-3-phosphoethanolamine (PE(17:0/17:0)), N-heptadecanoyl-D-*erythro*-sphingosylphosphorylcholine (SM(d18:1/17:0)), N-heptadecanoyl-D-*erythro*-sphingosine (Cer(d18:1/17:0)), 1,2-diheptadecanoyl-*sn*-glycero-3-phosphocholine (PC(17:0/17:0)), 1-heptadecanoyl-2-hydroxy-*sn*-glycero-3-phosphocholine (LPC(17:0)) and 1-palmitoyl-d31-2-oleoyl-*sn*-glycero-3-phosphocholine (PC(16:0/d31/18:1)), 1-palmitoyl-d31-2-oleoyl-*sn*-glycero-3-phosphocholine (PC(16:0/d31/18:1)) that were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL, USA), and tripalmitin-triheptadecanoylglycerol (TG(17:0/17:0/17:0)) (Larodan AB, Solna, Sweden). The samples were vortexed and incubated on ice for 30 min after which they were centrifuged ($9400 \times g$, 3 min, 4 °C). 60 μL from the lower layer of each sample was then transferred to a glass vial with an insert, and 60 μL of CHCl_3 : MeOH (2:1, v/v) was added to each sample. The samples were re-randomized and stored at -80 °C until analysis.

Calibration curves using 1-hexadecyl-2-(9Z-octadecenoyl)-*sn*-glycero-3-phosphocholine (PC(16:0/18:1(9Z))), 1-(1Z-octadecenyl)-2-(9Z-octadecenoyl)-*sn*-glycero-3-phosphocholine (PC(16:0/16:0)), 1-octadecanoyl-*sn*-glycero-3-phosphocholine (LPC(18:0)), (LPC18:1), PE (16:0/18:1), (2-aminoethoxy)[(2R)-3-hydroxy-2-[(11Z)-octadec-11-enoyloxy]propoxy]phosphinic acid (LysoPE (18:1)), N-(9Z-octadecenoyl)-sphinganine (Cer(d18:0/18:1(9Z))), 1-hexadecyl-2-(9Z-octadecenoyl)-*sn*-glycero-3-phosphoethanolamine (PE (16:0/18:1)) (Avanti Polar Lipids, Inc.), 1-Palmitoyl-2-Hydroxy-*sn*-Glycero-3-Phosphatidylcholine (LPC(16:0)) and 1,2,3 trihexadecanoalglycerol (TG16:0/16:0/16:0), 1,2,3-trioctadecanoylglycerol (TG(18:0/18:0/18:0)) and ChoE (18:0), 3 β -Hydroxy-5-cholestene 3-linoleate (ChoE(18:2)) (Larodan AB, Solna, Sweden), were prepared to the following concentration levels: 100, 500, 1000, 1500, 2000 and 2500 ng mL^{-1} (in CHCl_3 :MeOH, 2:1, v/v) including 1000 ng mL^{-1} of each internal standard.

The samples were analyzed using an established ultra-high-performance liquid chromatography quadrupole time-of-flight mass spectrometry method (UHPLC-QTOF-MS). The UHPLC system used in this work was a 1290 Infinity system from Agilent Technologies (Santa Clara, CA, USA). The system was equipped with a multi sampler (maintained at 10 °C), a quaternary solvent manager and a column thermostat (maintained

at 50 °C). Separations were performed on an ACQUITY UPLC® BEH C18 column (2.1 mm × 100 mm, particle size 1.7 µm) by Waters (Milford, USA).

The mass spectrometer coupled to the UHPLC was a 6545 quadrupole time of flight (QTOF) from Agilent Technologies interfaced with a dual jet stream electrospray ion (dual ESI) source. All analyses were performed in positive ion mode and MassHunter B.06.01 (Agilent Technologies) was used for all data acquisition. QC was performed throughout the dataset by including blanks, pure standard samples, extracted standard samples and control plasma samples. Relative standard deviations (% RSDs) for lipid standards representing each lipid class in the control plasma samples ($n = 8$) and in the pooled plasma samples ($n = 20$) were, on average, 11.7% (raw variation). The lipid concentrations in the pooled control samples was, on average, 8.4% and 11.4% in the standard samples. This shows that the method is reliable and reproducible throughout the sample set.

MS data processing was performed using the open-source software, MZmine 2.18(2). The following steps were applied in the processing: (1) Crop filtering with a m/z range of 350 – 1200 m/z and a RT range of 2.0 to 15.0 min, (2) Mass detection with a noise level of 1000, (3) Chromatogram builder with a min time span of 0.08 min, min height of 1200 and a m/z tolerance of 0.006 m/z or 10.0 ppm, (4) Chromatogram deconvolution using the local minimum search algorithm with a 70% chromatographic threshold, 0.05 min minimum RT range, 5% minimum relative height, 1200 minimum absolute height, a minimum ration of peak top/edge of 1.2 and a peak duration range of 0.08 - 5.0, (5) Isotopic peak grouper with a m/z tolerance of 5.0 ppm, RT tolerance of 0.05 min, maximum charge of 2 and with the most intense isotope set as the representative isotope, (6) Peak list row filter keeping only peaks with a minimum of 10 peaks in a row, (7) Join aligner with a m/z tolerance of 0.009 or 10.0 ppm and a weight for of 2, a RT tolerance of 0.1 min and a weight of 1 and with no requirement of charge state or ID and no comparison of isotope pattern, (8) Peak list row filter with a minimum of 53 peak in a row (10% of the samples), (9) Gap filling using the same RT and m/z range gap filler algorithm with an m/z tolerance of 0.009 m/z or 11.0 ppm, (10) Identification of lipids using a custom database search with an m/z tolerance of 0.009 m/z or 10.0 ppm and a RT tolerance of 0.1 min, (11) Normalization using internal standards (PE (17:0/17:0), SM (d18:1/17:0), Cer (d18:1/17:0), LPC (17:0), TG (17:0/17:0/17:0) and PC (16:0/d30/18:1)) for identified lipids and closest ISTD for the unknown lipids, followed by calculation of the concentrations based on lipid-class concentration curves, (12) Imputation of missing values by half of the row's minimum.

Sample selection

Figures S5 & S6 show intensities of the lipids across 121 samples obtained from 23 individuals who progressed to celiac disease (CD) and 112 samples from 23 matched healthy controls. The boxplots are sorted and color coded by age groups indicating the number of samples in each group from birth (cord blood, 0 month) until 84 months or 7 years of follow-up.

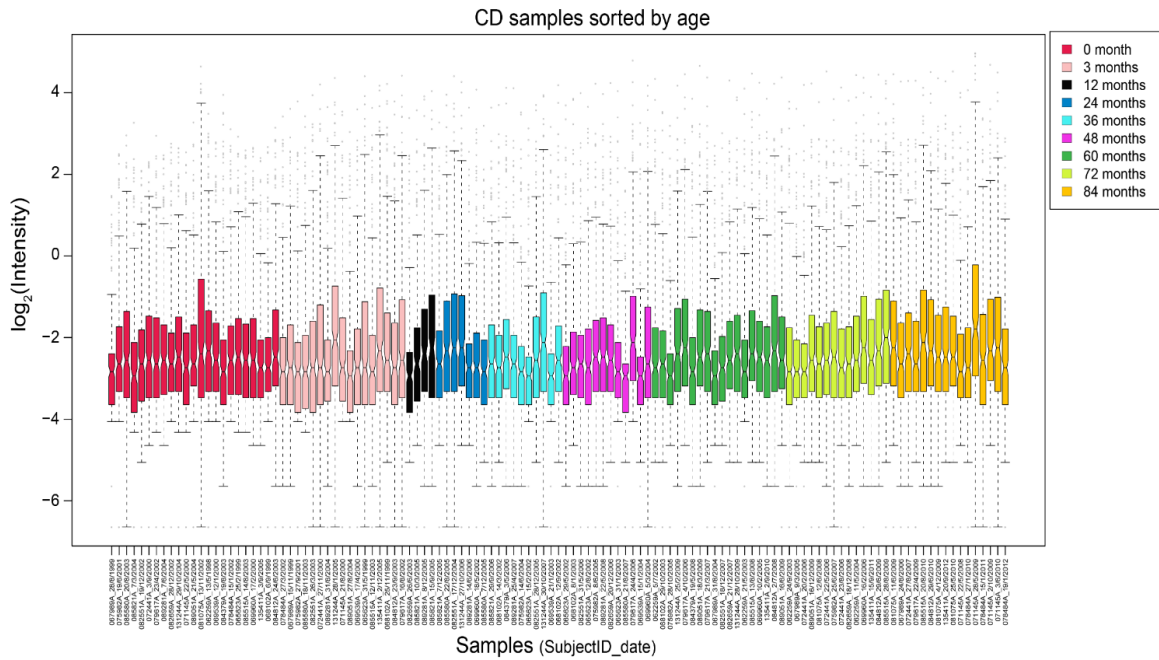


Figure S5. Log normalized intensities of the lipids measured in 121 plasma samples from 23 CD progressors. The samples are ordered by age and named in the pattern of 'subjectID_date'. The boxplot is colored by the age groups. The outliers are marked by grey dots. '0 month' denotes cord blood.

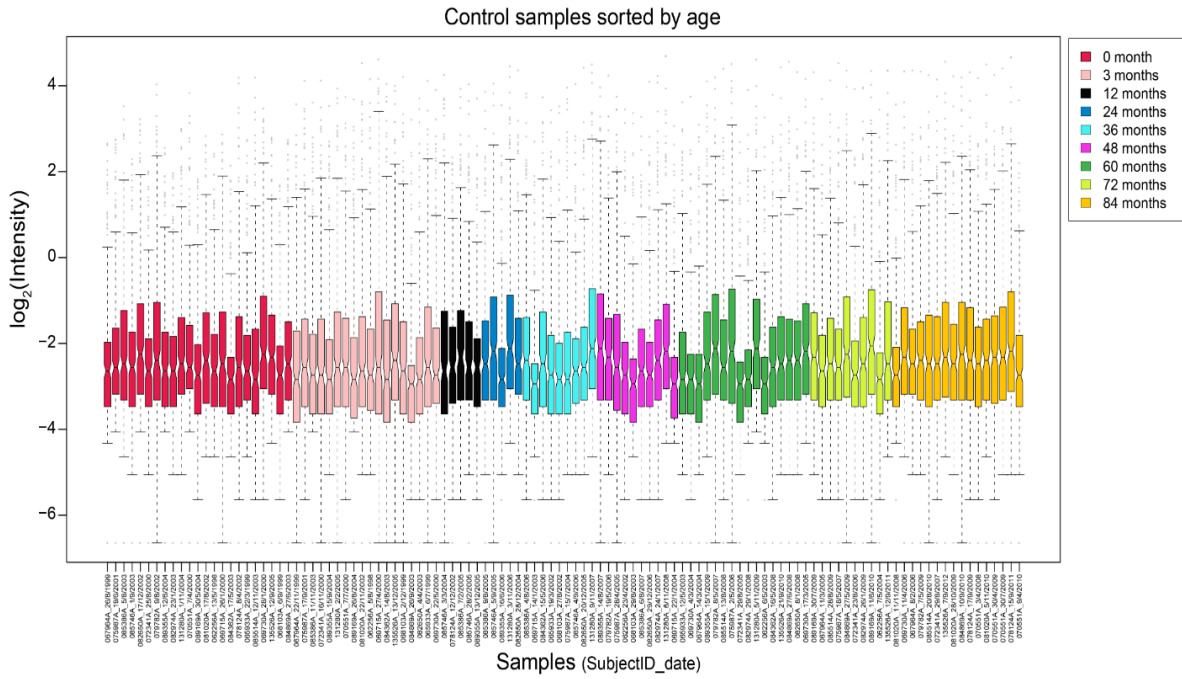


Figure S6. Log normalized intensities of the lipids measured in 112 plasma samples from 23 non-progressors (healthy controls). The samples are ordered by age and named in the pattern of 'subjectID_date'. The boxplot is colored by the age groups. The outliers are marked by grey dots. '0 month' denotes cord blood.

Distribution of subjects per age group

Distribution of the samples collected in this study from healthy children and those who later progressed to CD were sorted and grouped according to the follow-up age as shown in **Figure S7**.

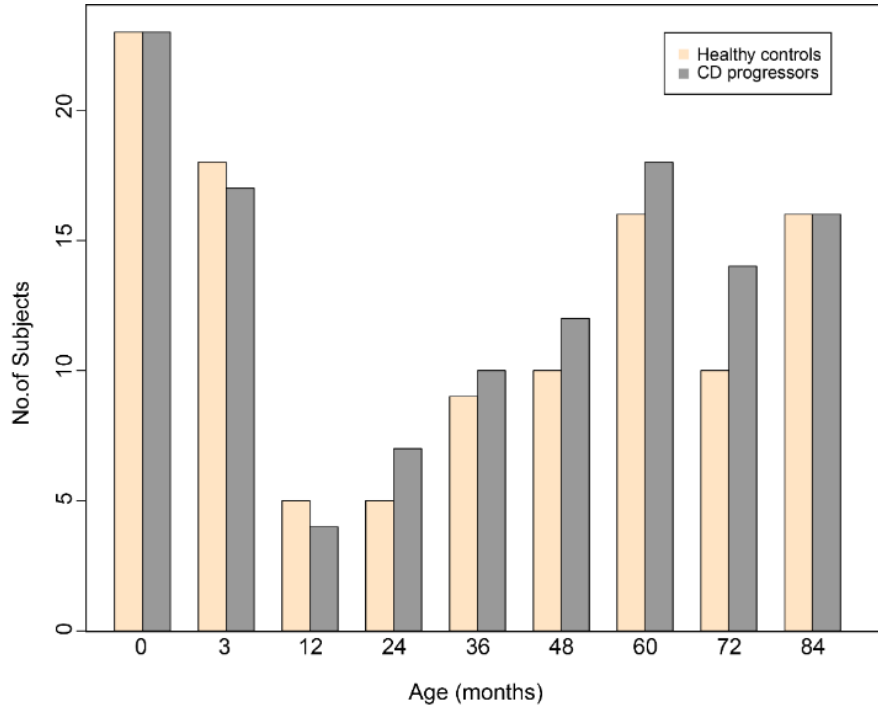


Figure S7. Bar plot showing number of study samples obtained from CD progressors (grey) and healthy controls (orange) per age group.

Factors and sources of variation in lipidomics dataset

Effects of different factors such as age, gender and status (healthy or CD) on lipidomics data were evaluated. The data were centered to zero mean and unit variance. The relative contribution of each factor (experimental variable) to the total variance in the dataset was estimated by fitting the linear model regression model, where the normalized intensities of metabolites regressed to the factor of interest, and thereby estimating the median marginal coefficients (R^2). This analysis was performed using package 'scater'.

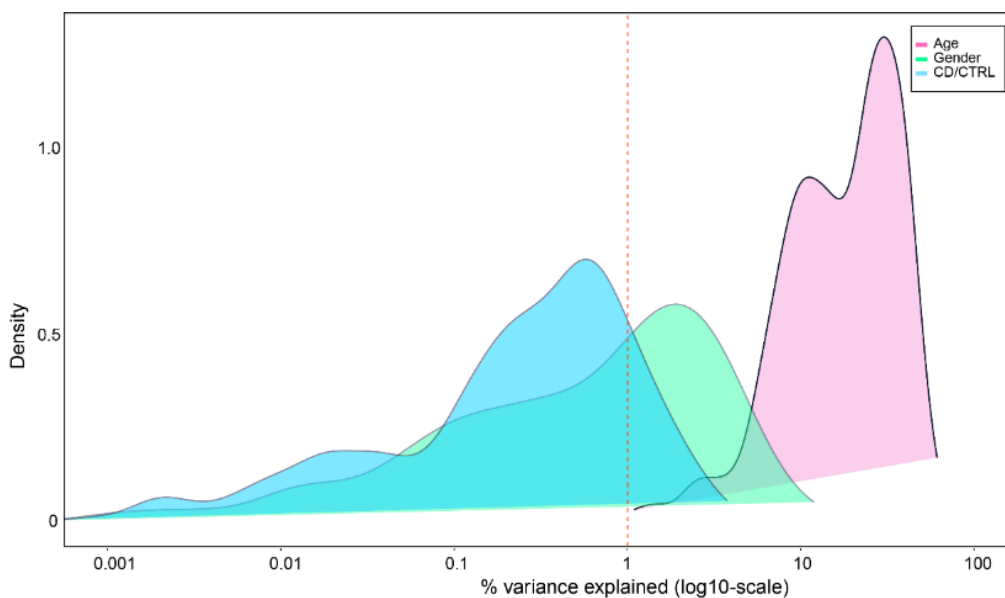


Figure S8. Display a density plot of the metabolite-wise marginal R^2 values for the factors.

Lipidome analysis of healthy controls using PLS-DA (supervised)

PLS-DA(3, 4) model for healthy controls developed from 112 follow-up samples showed difference in lipidome along the follow-up age. As in the CD progressors (**Figure 1C**), the cord blood lipidome was distinctly separated from other age groups (**Figure S9**). We have not found any significant differences in the cord blood lipids between CD progressors and healthy controls (**Figure 2A**). However, some lipids changed significantly at 3 months between progressors and controls. These changes exacerbated at a later age (**Figure 2A & Figure 3**).

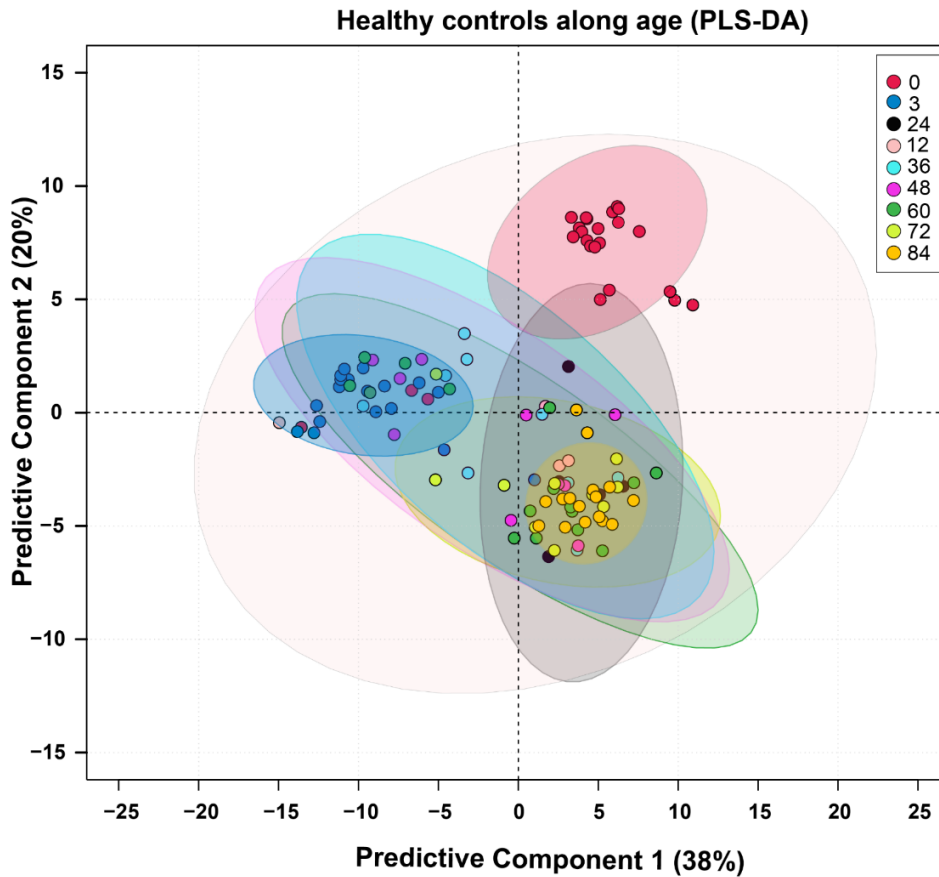


Figure S9. Score plot of PLS-DA model ($Q^2 = 0.514$ and $R^2Y=0.685$) showing similarity or difference in the lipidome of healthy control from birth. The groups are labeled as age in months.

Change in total TG, PC and SM after gluten intake

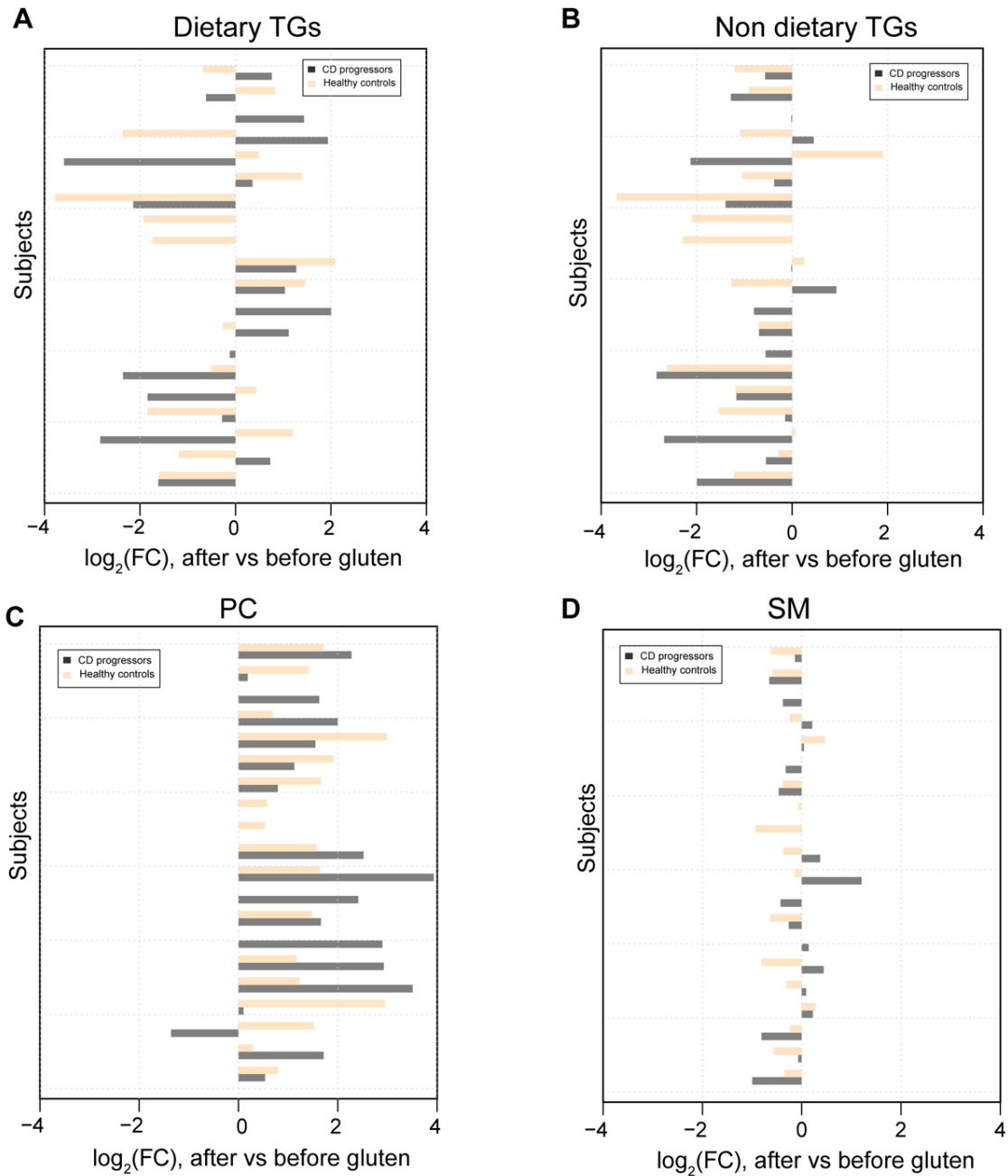


Figure S10 Bar plots showing log₂ fold changes (after vs. before introduction of gluten) of total triglycerides (TGs), phosphatidylcholine (PCs), sphingomyelins (SMs) in individual CD progressor (grey) and their matched healthy control (orange).

Monitoring the celiac disease progression

Plasma samples were collected from the subjects who progressed to CD at a particular stage i.e., cord blood (not shown here), after 3 months of birth, before and after seroconversion (tTGA detected), endoscopy/introduction of GFD and 6 to 12 months after GFD has been recommended. Subject's ages (range) along the spectrum of diagnosis and treatment is shown in **Figure S11**.

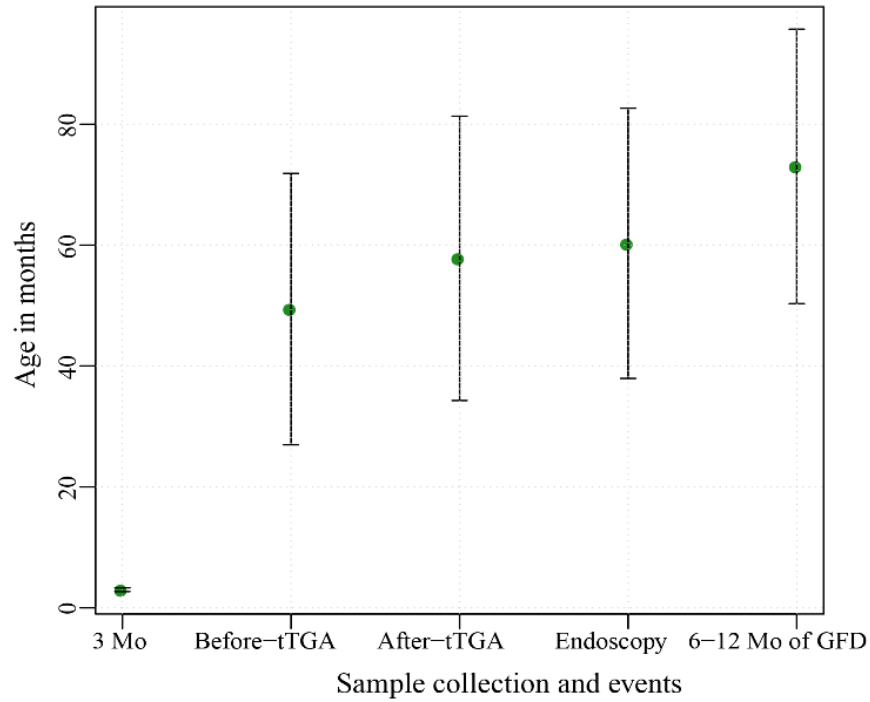


Figure S11

References

1. Folch J, Lees M, Sloane Stanley G. A simple method for the isolation and purification of total lipids from animal tissues. *J Biol Chem* 1957;226:497-509.
2. Pluskal T, Castillo S, Villar-Briones A, et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform* 2010;11:395.
3. Pinheiro J, Bates D, DebRoy S, et al. Linear and nonlinear mixed effects models. R package version 3 2014.
4. Pinheiro JC, Bates DM. *Mixed-effects models in S and S-PLUS* Springer. New York 2000.