

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

Partho Sen¹, Cecilia Carlsson², Suvi M. Virtanen^{3,4,5,6}, Satu Simell⁷, Heikki Hyöty^{8,9}, Jorma Ilonen^{10,11}, Jorma Toppari^{7,12}, Riitta Veijola^{13,14,15}, Tuulia Hyötyläinen², Mikael Knip^{16,17*}, Matej Orešić^{1,18*}

¹Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, FI-20520 Turku, Finland

²Department of Chemistry, Örebro University, 702 81 Örebro, Sweden

³National Institute for Health and Welfare, Public Health Promotion Unit, Helsinki, Helsinki, Finland

⁴Faculty of Social Sciences/Health, University of Tampere, Tampere, Finland

⁵Center for Child Health Research, University of Tampere and Tampere University Hospital, Tampere, Finland

⁶Science Centre, Tampere University Hospital, Tampere, Finland

⁷Department of Paediatrics and Adolescent Medicine, Turku University Hospital, Turku, Finland

⁸Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland

⁹Fimlab Laboratories, Pirkanmaa Hospital District, Tampere, Finland

¹⁰Immunogenetics Laboratory, Institute of Biomedicine, University of Turku, Turku, Finland

¹¹Clinical Microbiology, Turku University Hospital, Turku, Finland

¹²Institute of Biomedicine, Centre for Integrative Physiology and Pharmacology, University of Turku, Turku, Finland

¹³Department of Paediatrics, PEDEGO Research Unit, Medical Research Centre, University of Oulu, Oulu, Finland

¹⁴Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland

¹⁵Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

¹⁶Children's Hospital, University of Helsinki and Helsinki University Hospital, 00290 Helsinki, Finland; Research Programs Unit, Diabetes and Obesity, University of Helsinki, 00290 Helsinki, Finland

¹⁷Tampere Centre for Child Health Research, Tampere University Hospital, Tampere, Finland

¹⁸School of Medical Sciences, Örebro University, 702 81 Örebro, Sweden

*Corresponding authors:

Email: mikael.knip@helsinki.fi or matej.oresic@utu.fi

ABSTRACT

Background and Aims

Celiac disease (CD) is a chronic enteropathy characterized by an autoimmune reaction in the small intestine in genetically-susceptible individuals. Gluten is the required environmental trigger of clinical CD, but the underlying causes of the autoimmune reaction remain unknown. Herein, we apply lipidomics to elucidate the early events preceding clinical CD in a prospective study of children observed from birth until diagnosis of CD and subsequent introduction of a gluten-free diet.

Methods

Mass spectrometry-based lipidomics profiling was applied to a longitudinal series of 233 plasma samples from the Type 1 Diabetes Prediction and Prevention (DIPP) study, spanning the period between birth and the introduction of a gluten-free diet following CD diagnosis (n=23 CD progressors, n=23 controls matched for gender, HLA risk, period of birth, and age).

Results

23 children progressed to CD at a mean age of 4.8 years. They showed increased amounts of triacylglycerols (TGs) of low carbon number and double bond count and a decreased level of phosphatidylcholines by 3 months of age as compared to controls. These differences were exacerbated with age but were not observed at birth. No significant differences were observed in essential (dietary) TGs such as those containing polyunsaturated fatty acids.

Conclusion

Our findings suggest that abnormal lipid metabolism associated with development of clinical CD may occur prior to the introduction of gluten to the diet. Moreover, our data suggest that the specific TGs found elevated in CD progressors may be due to a host response to compromised intake of essential lipids in the small intestine, requiring *de novo* lipogenesis.

Keywords: celiac disease, lipid metabolism, lipidomics, metabolomics, triacylglycerols

INTRODUCTION

Celiac disease (CD) is a chronic, systemic, autoimmune enteropathy triggered by dietary gluten and related prolamines from rye and barley in genetically-susceptible individuals¹⁻⁶. About 90-95% of CD patients express HLA-DQ2 protein, while the remaining (5-10%) express HLA-DQ8^{7, 8}. CD is characterized by a wide range of gastro- and extraintestinal symptoms that include diarrhea, weight loss, abdominal distention, malabsorption and iron-deficiency anemia^{1, 4, 9}. Serologic tests such as measurement of serum IgA and/or IgG tissue transglutaminase antibodies, IgA endomysial antibodies, deamidated gliadin peptide antibodies (IgG class) are performed for screening and diagnosis of CD. In addition, a biopsy of small intestine is still required in many countries to confirm the diagnosis^{2, 4}.

The incidence of CD and other autoimmune diseases such as type 1 diabetes (T1D) have been increasing in children and adults over the past decades^{10, 11}. The occurrence of these autoimmune diseases is higher in the Nordic countries¹² than elsewhere, with the highest prevalence of CD occurring in Sweden (29/1000 by the age of 12) and the highest incidence rate of T1D occurs in Finland (64/100000/year for children under 15 years of age)^{13, 14}. Approximately 10% of patients with T1D develop overt CD^{15, 16}. On the other hand, people with CD are at-risk for T1D before 20 years of age¹⁷. These autoimmune diseases share common, predisposing alleles in the class II HLA-region as the DR3-DQ2 and DR4-DQ8 haplotypes¹⁸⁻²¹.

Recent studies show that, in addition to genetic predisposition and exposure to dietary gluten, other factors such as the composition of the intestinal microbiota, birth delivery mode, infant feeding and the use of antibiotics may also affect the onset of CD^{1, 22}. Thus, the early pathogenesis of CD is still poorly understood^{4, 23, 24} and the identification of molecular signatures associated with progression to overt CD remains an unmet medical need^{3, 4, 24, 25}. The health burden of CD in terms of quality of life, complications, mortality and cost of treatment are considerable, meaning that its prevention has become, in the last decade, an important area of research.

Metabolomics is the study of small (< 1500 Dalton) molecules and their functions in cells, tissues and body fluids²⁶. Metabolomic studies in adults diagnosed with active CD²⁷ identified marked changes in serum and urine metabolic profiles, along with altered intestinal microbiota²⁸⁻³⁰. Interestingly, a recent, small prospective study suggests an altered early trajectory of the gut microbiome (4 and 6 months of age) in children who later progressed to CD³¹. Given the characteristic metabolic abnormalities observed in clinical CD, we investigate here whether distinct lipidomic signatures exist prior to overt CD. We analyzed molecular lipids in plasma by lipidomics approach in a prospective cohort of children, observed from birth until clinical CD.

MATERIALS AND METHODS

Study design and protocol

The children included in the present study are from the DIPP cohort, which is an ongoing prospective study initiated in 1994. In the DIPP study, parents of newborn infants at the university hospitals of Turku, Tampere and Oulu in Finland are asked for permission to screen the child for HLA alleles conferring risk of T1D, using umbilical cord blood. Families of children identified as having an increased HLA-conferred risk for T1D are invited to join the study. Our analysis included children born at Tampere University Hospital between August 1999 and September 2005. During that period, 23,839 children were screened at birth for increased risk of T1D, and 2,642 eligible children were enrolled in the follow-up and had at least two visits to the study clinic. These children carry the high-risk HLA DQB1*02/*03:02 genotype or the moderate-risk HLA-DQB1*03:02/x genotype (x ≠ DQB1*02, 03:01, or 06:02). More than 1,200 of these children took part in the DIPP-CD study. The DIPP study in Tampere followed the children at regular intervals at the ages of 3, 6, 12, 18, and 24 months and subsequently at intervals of 12 months for children without T1D-related autoantibodies. At each visit, the families were interviewed for diet, infections, growth, important family-related issues and the children gave a non-fasting venous blood sample. The children were followed for four CD-related antibodies: anti-tissue transglutaminase (anti-tTG), anti-endomysium (EMA), antigliadin (AgA-IgG and AgA-IgA) and anti-reticulin (ARA) antibodies, and for T1D-associated autoantibodies: islet cell antibodies (ICA). If a sample was positive for ICA, then insulin autoantibodies (IAA),

antibodies against tyrosine phosphatase-like protein (IA-2A) and glutamate decarboxylase (GADA) were measured from all samples taken from that child. From the beginning of 2003, all samples were measured for the four T1D-associated autoantibodies.

All children participating in this study were of Caucasian origin. IgA deficiency was excluded. One of the mothers had CD. The children in the CD follow-up cohort were annually screened for anti-tTG antibodies (tTGA) using a commercial kit (Celikey Pharmacia Diagnostics, Freiburg, Germany). If any given child's sample was found to be positive for tTGA, all of that child's previous and following samples were analyzed for the entire set of CD-related antibodies. A duodenal biopsy was recommended for all tTGA-positive children. If the biopsy was consistent with the ESPGHAN criteria of 1990, a gluten-free diet (GFD) was recommended. Three of the CD cases were also diagnosed with T1D, one just before the diagnosis of CD and the other ones after the diagnoses of CD (62 and 76 months later).

We randomly selected 23 children (12 males and 11 females) with biopsy-proven CD (progressors) and a control for each progressor matched for age, gender, and with the same risk HLA alleles, and living in the Tampere region throughout the whole follow-up period. None of the controls were diagnosed with T1D. The original HLA screening for CD patients and suitable controls was also completed to a “full house” DQ typing to verify the presence of DQA1*05 in selected DQB1*02 positive children³². These clinical and genetic annotations of the participants are given in **Table 1**. Furthermore, the children were exclusively breastfed until 1.6 months (median) and some were exclusively breastfed up to 8.0 months of age. The first exposure to gluten in this study was at a median age of 6.0 months. The children (case and control) had comparable (nominal, non-significant difference) energy (KJ), fat (g), carbohydrate (oat, rye, wheat), and gluten intakes (see **Supplementary Figures S1 & S2**).

Maternal total diet during pregnancy and lactation was ascertained by a validated food frequency questionnaire³³. Infant feeding was studied by structured questionnaires, which families filled in at home, and which were checked during the clinic visits. The child's diet

was assessed with 3-day food records at 3, 6, and 12 months followed by 2, 3, 4 and 6 years of age visits. The food records were posted before the study visit or given at the previous visit with detailed, written instructions. The families and daycare personnel were instructed to write down all the foods, drinks and dietary supplements the child consumed, with the amounts and brand names during one weekend day and two week-days. The study nurses checked the records at the study visit and helped to estimate any incomplete portion sizes with a picture booklet. The collection, processing and calculation of food consumption data has been described previously in detail³⁴. The mothers of CD progressors and controls had comparable body mass indices (BMIs) and energy, fat and carbohydrate intakes during pregnancy and lactation (**Supplementary Figures S3 & S4**).

Altogether, 121 plasma samples from children developing CD and 112 plasma samples from matched, healthy controls were analyzed, averaging more than five prospective follow-up samples for each child. Log-transformed intensities of the lipid measurements from these are given in **Supplementary Figures S5 & S6**.

Analysis of molecular lipids

A total of 233 plasma samples were randomized and extracted using a modified version of the Folch procedure³⁵. Promptly after extraction, 10 µL of 0.9% NaCl and 120 µL of CHCl₃: MeOH (2:1, v/v) containing 2.5 µg 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 × 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 CHCl₃: 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), 3β-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 CHCl₃: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 2.18³⁶. 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.

Statistical methods

The samples from CD progressors and controls were divided into different age groups based on time difference between the date of the sample withdrawn and date of birth of the subject (**Supplementary Figure S7**). If more than two samples from the same case

matched a time interval, the closest was taken. The data was then log₂ transformed. Homogeneity of the samples were assessed by principal component analysis (PCA)³⁷ and no outliers were detected.

The R software (<http://www.r-project.org>) was used for data analysis and visualization. PCA was performed using ‘*prcomp()*’ function included in the ‘*stats*’ package. The effect of different factors such as age, gender and condition (healthy or CD) on the lipidomics dataset was evaluated. The data was centered to zero mean and unit variance. The relative contribution of each factor to the total variance in the dataset was estimated by fitting a linear regression model, where the normalized intensities of metabolites were regressed to the factor of interest, and thereby median marginal coefficients (R^2) were estimated. This analysis was performed using ‘*scater*’ package (**Supplementary Figure S8**). Partial least squares Discriminant Analysis (PLS-DA)³⁸⁻⁴⁰. and Variable Importance in Projection (VIP) scores⁴¹ were estimated by an array of functions coded in ‘*ropels*’ package. Moreover, the PLS-DA models were cross-validated⁴² by 7 fold cross-validation as implemented in ‘*ropels*’ package and Q^2 (an estimate of model’s predictability) and R^2 were obtained.

The longitudinal profiles of the lipids in the samples obtained from CD progressors and matched healthy controls were compared using linear mixed-effects (LME) models^{43, 44} as implemented in the ‘*lme()*’ function of ‘*nlme*’ package. The intensity of a metabolite (\hat{y}) in a sample (j) is a function of multiple factors such as follow-up age, gender, case/control, subject-wise variability, etc. The clinical and genetic factors of the children were matched and standardized in this study (**Table 1**), other factors such as subject-wise variability can be modelled and derived using LME models. The LME models were restricted to constant terms with the fixed effect being CD progression/non-progression (control), age, gender and the random effect being subject-wise variation in comparison to the group-specific mean level. Fitted LME models showed, gender difference has nominal effect on the metabolite intensities. The fully parametrized model was compared to a null model using analysis of variance (ANOVA)⁴⁵ (‘*aov()*’ as deployed in the ‘*stats*’ package). The lipid profiles that changed significantly ($p < 0.05$) were subjected to False Discovery Rates (FDR) adjustment using *p-adjust()*. Lipid profiles with $FDR < 0.05$ were listed.

Next, post-hoc analysis using Tukey's test for Honest Significant Difference (HSD) was performed on these selected lipids to see, if at all, they are changing between progressors and non-progressors at a particular age. A list of differentially altered lipids that also showed HSD (p -values < 0.05) between CD progressor and healthy control at a particular age was marked. These lipids were considered for further analysis.

HSD was performed using '*TukeyHSD()*' function deployed in the 'stats' package. Loess regression was performed using '*loess()*' deployed in the 'stats' package. Spearman's correlation coefficient was calculated using '*rcorr()*' function implemented in '*Hmisc*' package. '*Heatmap.2()*' and '*boxplot()*' was used for data visualization.

RESULTS

Global plasma lipidome in progression to celiac disease

The complete lipidomics dataset was first explored using multivariate analysis. Among other factors affecting the lipidome, age was found to be a major confounding factor (**Supplementary Figure S8**). Partial least square discriminant analysis (PLS-DA)³⁸⁻⁴⁰ of lipidomics data from all 233 longitudinal samples suggested that children who later progressed to CD (progressors) may have different lipid profiles in comparison to their matched, healthy controls (**Figure 1a**). Different classes of lipids such as cholesterol esters (CEs), phosphatidylcholines (PCs), lysophosphatidylcholines (LPCs), phosphatidylethanolamines (PEs), phosphatidylinositols (PIs), sphingomyelins (SMs) and triacylglycerols (TGs) were affected (regression coefficient, RC (± 0.05) and VIP scores⁴¹ > 1) (**Figure 1b**).

Differences between cases and controls were observed already by 3 months of age, *i.e.*, before the introduction of gluten to the diet (**Figure 1c**). TGs, PCs and CEs were mostly affected (RC (± 0.05) and VIP⁴¹ scores > 1) at this age (**Figure 1d**). The cord plasma lipidome clustered distinctly from other age groups (**Supplementary Figure S9**). However, no significant differences (HSD, p -values > 0.05) in the cord plasma lipids were observed between CD progressors and their matched healthy controls (**Figure 2a**).

Molecular lipids in progression to overt CD

Longitudinal analysis of lipidomic profiles at the individual lipid level identified 80 molecular lipids (from a total of 239) occurring at significantly-different levels over time between CD progressors and healthy controls (FDR-adjusted p-values < 0.05) (**Figure 2a**).

The level and direction of regulation of TGs depends on their chemical structure. There was an association between the fold change in TGs (CD progressors vs. healthy controls) and the TG double bond count as well as the TG carbon number (**Figure 2b-e**). At 3 months of age, TGs with lower double bond and carbon numbers were found to have increased in CD progressors as compared to matched, healthy controls, with coefficients ($R = -0.13$; p -values = 0.19) and ($R = -0.54$; p -values = 0.0001) respectively for double bond count and carbon number (**Figures 2b** and **2d**). Interestingly, at a later age, these TGs with lower double bond and carbon numbers were down-regulated with ($R = 0.79$; p -values = 0.08) and ($R = 0.44$; p -values = 0.0002), respectively (**Figure 2c** and **Figure 2e**). Notably, four such TGs were elevated in progressors at 3 months of age (HSD, p -values < 0.05). No significant changes were detected for dietary TGs; such as those containing polyunsaturated fatty acids (PUFAs). Representative longitudinal profiles for selected significantly-altered lipids between CD progressors and healthy controls are shown in **Figure 3**.

Impact of tissue transglutaminase antibodies on lipidomic profiles

Tissue transglutaminase (tTG) is a multifunctional, calcium-dependent enzyme (EC 2.3.2.13) that plays an important role in the pathogenesis of CD. Antibodies produced against tTG, anti-tissue transglutaminase antibodies (tTGA), are used as serological markers with high sensitivity (99%) and reasonable specificity (> 90%) for the diagnosis of CD^{2, 46, 47}.

We examined tTGA titers in 23 CD progressors at two different stages, (1) immediately after the seroconversion for tTGA, and (2) 6-12 months after both seroconversion and introduction of GFD. First, we identified at least seven ‘essential’ TGs, *i.e.*, TGs of dietary origin, in our lipidomics dataset. The identities of these TGs were validated using external sources including both the Human Metabolome Database (HMDB)⁴⁸ and FooDB.ca

(<http://foodb.ca/>). Spearman correlation analysis was then performed between the levels of selected TGs (34 significantly-changed, nonessential and seven non-significantly changed essential TGs) and PCs (18 PCs + 2 LPCs) with measured tTGA levels at these stages. The aim was to determine whether the systemic levels of TGs and PCs are associated with antibody titers during CD progression. A positive correlation (Spearman's rank coefficient, $\delta_{avg} = +0.174$; minimum p-value = 0.01) was found between the TG and tTGA levels after seroconversion (**Figure 4a**). However, a negative correlation was found at a later age, where the levels of tTGA decreased considerably ($\delta_{avg} = -0.163$; minimum p-value = 0.05). In addition, at least 16 PCs were negatively correlated at these two different time-points ($\delta_{avg} = -0.02$ and -0.13 ; minimum p-values = 0.01 and 0.02 respectively) (**Figure 4b**). It is known that GFD decreases tTGA levels in CD patients⁴⁹, whilst it also improves the total cholesterol levels and high density lipoprotein (HDL) profiles without any significant increase in low density lipoproteins (LDL)⁵⁰.

Impact of gluten on molecular lipids

We then divided the cohort into three subgroups, (1) before the introduction of gluten in the diet (3 months) (2) after the introduction of gluten (12-36 months), and (3) after the introduction of GFD (72 months, CD progressors only).

We observed a decrease in total essential TG level in the plasma of the CD progressors after gluten intake (**Figure 5a**, **Supplementary Figure S10a**). Introduction of GFD reversed this trend, but the changes were not significant (ANOVA, p-values > 0.05) (**Figure 5a**, **Figure 5e**). On the other hand, levels of nonessential endogenous TGs were decreased (ANOVA, p-values = 0.004) in the plasma of the CD progressors after commencement of gluten intake and even after the diagnosis of clinical CD and introduction of GFD (p-values = 0.001). The trend in the change of nonessential TGs was also observed in healthy controls (p-values = 0.007 and 0.01, respectively) (**Figure 5b**, **Figure 5e** and **Supplementary Figure S10b**). Moreover, there was a difference (p-values < 0.05) in nonessential TGs levels between CD progressors and healthy controls as predicted by Tukey's test for HSD (**Figure 2a**). These findings suggest that dysregulation of lipid metabolism might occur at an early stage of CD progression, even before the introduction of gluten in the diet.

PCs were elevated both in the progressors ($p < 0.004$) and controls ($p < 0.007$) after the commencement of gluten intake (**Figure 5c**, **Figure 5e** and **Supplementary Figure S1oc**). PCs are the major class of phospholipids that form the cellular constituents required for the assembly of biological membranes. In contrast to healthy controls, no significant differences in SM level were observed in the CD progressors after commencement of gluten intake. A difference in the plasma SM level was observed only at a later age, after the introduction of GFD (**Figure 5d**, **Figure 5e** and **Supplementary Figure S1od**).

DISCUSSION

We identified systematic differences in plasma lipidomes between children who progressed to clinical CD as compared to children who remained healthy during the follow-up. These differences were observed before the exposure to gluten in the diet and before the first signs of CD-associated autoimmunity. The dysregulation of the plasma lipidome in CD progressors is dominated by complex lipids such as PLs, TGs and CEs. As there is no evidence of gut damage in any studies performed so far in individuals who are autoantibody-negative for tTGA, it is unlikely that these systematic differences are caused by gluten from breast milk, or other sources in the infants' diet causing damage to the gut immediately after birth. An earlier lipidomics study in the PreventCD cohort did not find significant differences between CD progressors and matched controls at 4 months of age⁵¹. However, that study used a targeted analysis focusing on a subset of phospholipids and acyl-carnitines, and did not measure TGs and CEs, where the major changes were found to have occurred in our study.

In this study, several nonessential TGs were up-regulated in CD progressors as early as 3 months of age, *i.e.*, before the first introduction of gluten to the diet. However, no significant changes in dietary TGs were found at this age. Lipid malabsorption is believed to be a side effect of flattened villi in the small intestine^{27, 29}. However, our data suggest that lipid-related abnormalities are not caused by CD-related villous atrophy in the gut. One plausible explanation that arises is that lipid malabsorption can lead to a reduction in cholesterol-transporting lipoproteins, and secretion of apolipoprotein (Apo)-A1^{52, 53}. A

decrease in the level of CEs in CD progressors during the first 3 months after birth supports this explanation. Moreover, CEs were up-regulated at a later age after the introduction of GFD (**Figure 2a** and **Figure 3a**). Lipid malabsorption may thus occur at a very early age in CD progressors and, therefore, *de novo* lipogenesis⁵⁴, as reflected by increased levels of TGs with low carbon number and double bond content, may be necessarily increased in order to compensate for said compromised lipid uptake. Interestingly, this phenomenon was not observed at birth (*i.e.*, in the cord blood data), thus suggesting that the observed lipid dysregulation is not an inborn phenomenon.

Furthermore, an inverse relationship between endogenous TGs and tTGA titer after seroconversion and introduction of GFD in the CD progressors also reaffirms the involvement of TGs and other phospholipids (PCs) in CD progression. Indeed, there was an increase in cholesterol levels after the introduction of GFD and there was a decrease in tTGA titers, in agreement with previous studies^{49, 50}. Presumably, these lipids may play a protective role against seroconversion to tTGA positivity and CD progression.

We found, in earlier studies, that T1D is preceded by dysregulation of lipid metabolism⁵⁵⁻⁵⁷. The key findings from these studies were that islet autoimmunity and overt T1D are preceded by diminished phospholipid and TG levels. While there are some similarities at an early age in phospholipid profiles in T1D and, as shown in the current study, CD progressors, no such similarities exist for TGs. The increase of TGs with low carbon number and double bond count does appear to be specific to CD progression. These TGs are associated, in adults, with elevated liver fat in non-alcoholic fatty liver disease (NAFLD)^{58, 59}, reflecting increased *de novo* lipogenesis and adipose tissue lipolysis. In fact, CD patients have, interestingly, been found to be at increased risk of NAFLD⁶⁰. Our findings may thus offer an explanation for this association.

The main limitation of this study is the relatively small number of children included. However, this is mitigated by the longitudinal study setting, with five prospective samples, from birth until after the introduction of GFD, on average, from each child. The subjects of our study were well-matched with their controls for genetic and environmental factors.

In summary, our study suggests that lipid-related abnormalities in CD progressors demonstrably occur prior to the first introduction of gluten to the diet. These changes may be related to impaired lipid absorption as well as *de novo* lipogenesis, with both of these being part of the host response. Our findings may, therefore, have important clinical implications for the detection of subjects at-risk for CD as well as for the understanding of early pathogenesis of CD.

DATA ACCESSIBILITY

The lipidomics datasets and the clinical metadata generated in this study have been submitted to MetaboLights⁶¹ and can be located using accession number (*MTBLS729*). The appropriate clinical metadata was linked to the lipidomics dataset using the ISA-creator package from MetaboLights.

ETHICAL APPROVAL AND INFORMED CONSENT

The ethics committee of Tampere University Hospital approved the study. Written, informed consent was obtained from the parents for HLA-screening, autoantibody analysis and intestinal biopsies.

ACKNOWLEDGEMENTS

We thank the families participating in the DIPP study for making this study possible. We also thank the expert staff of the DIPP study for their excellent work with the participating research families and sample collection. We thank Professor Olli Simell for his important scientific contribution to the DIPP study. We thank to Dawei Geng for technical assistance in lipidomic analysis and to Aidan McGlinchey for editing the manuscript.

COMPETING INTERESTS

None declared.

FUNDING

This study was supported by the Academy of Finland (Centre of Excellence in Molecular Systems Immunology and Physiology Research – SyMMyS, Decision No. 250114, to M.O. and M.K.; and Personalised Health 2014 programme project, Decision No. 292568).

REFERENCES

1. Kemppainen KM, Lynch KF, Liu E, et al. Factors that increase risk of celiac disease autoimmunity after a gastrointestinal infection in early life. *Clin Gastroenterol Hepatol* 2017;15:694-702. e5.
2. Oxentenko AS, Murray JA. Celiac disease: ten things that every gastroenterologist should know. *Clin Gastroenterol Hepatol* 2015;13:1396-1404.
3. Schuppan D, Junker Y, Barisani D. Celiac disease: from pathogenesis to novel therapies. *Gastroenterology* 2009;137:1912-33.
4. Kelly CP, Bai JC, Liu E, et al. Advances in diagnosis and management of celiac disease. *Gastroenterology* 2015;148:1175-86.
5. Koning F, Gilissen L, Wijmenga C. Gluten: a two-edged sword. Immunopathogenesis of celiac disease. Springer Semin Immunopathol 2005;27:217-32.
6. Vader W, Kooy Y, van Veelen P, et al. The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 2002;122:1729-1737.
7. Stankovic B, Radlovic N, Lekovic Z, et al. HLA genotyping in pediatric celiac disease patients. *Bosn J Basic Med Sci* 2014;14:171-6.
8. Romanos J, Van Diemen CC, Nolte IM, et al. Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology* 2009;137:834-840. e3.
9. Disease RcfDoC. Report of working group of European society of paediatric gastroenterology and nutrition. *Arch Dis Child* 1990;65:909-911.
10. Mustalahti K, Catassi C, Reunanan A, et al. The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. *Ann Med* 2010;42:587-595.
11. Patterson CC, Dahlquist GG, Gyurus E, et al. Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet* 2009;373:2027-33.
12. Rewers M. Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? *Gastroenterology* 2005;128:S47-S51.
13. Myléus A, Ivarsson A, Webb C, et al. Celiac disease revealed in 3% of Swedish 12-year-olds born during an epidemic. *J Pediatr Gastroenterol Nutr* 2009;49:170-176.
14. Knip M. Descriptive epidemiology of type 1 diabetes—is it still in? *Diabetologia* 2012;55:1227-1230.
15. Rewers M, Liu E, Simmons J, et al. Celiac disease associated with type 1 diabetes mellitus. *Endocrinol Metab Clin North Am* 2004;33:197-214, xi.
16. Scaramuzza AE, Mantegazza C, Bosetti A, et al. Type 1 diabetes and celiac disease: The effects of gluten free diet on metabolic control. *World J Diabetes* 2013;4:130-4.
17. Ludvigsson JF, Ludvigsson J, Ekbom A, et al. Celiac disease and risk of subsequent type 1 diabetes: a general population cohort study of children and adolescents. *Diabetes Care* 2006;29:2483-8.
18. Bakker SF, Tushuizen ME, von Blomberg BM, et al. Screening for coeliac disease in adult patients with type 1 diabetes mellitus: myths, facts and controversy. *Diabetol Metab Syndr* 2016;8:51.

19. Holmes GK. Screening for coeliac disease in type 1 diabetes. *Arch Dis Child* 2002;87:495-8.
20. Ludvigsson JF, Lebwohl B. Screening for coeliac disease in type 1 diabetes. *Acta Paediatr* 2017;106:530-531.
21. Smyth DJ, Plagnol V, Walker NM, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med* 2008;359:2767-2777.
22. Lionetti E, Catassi C. The Role of Environmental Factors in the Development of Celiac Disease: What Is New? *Diseases* 2015;3:282-293.
23. Godkin A, Jewell D. The pathogenesis of celiac disease. *Gastroenterology* 1998;115:206-10.
24. Bernini P, Bertini I, Calabro A, et al. Are patients with potential celiac disease really potential? The answer of metabonomics. *J Proteome Res* 2011;10:714-21.
25. Scanlon SA, Murray JA. Update on celiac disease - etiology, differential diagnosis, drug targets, and management advances. *Clin Exp Gastroenterol* 2011;4:297-311.
26. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. *Proteomics* 2006;6:4716-23.
27. Solakivi T, Kaukinen K, Kunnas T, et al. Serum fatty acid profile in celiac disease patients before and after a gluten-free diet. *Scand J Gastroenterol* 2009;44:826-30.
28. Di Cagno R, De Angelis M, De Pasquale I, et al. Duodenal and faecal microbiota of celiac children: molecular, phenotype and metabolome characterization. *BMC Microbiol* 2011;11:219.
29. Bertini I, Calabro A, De Carli V, et al. The metabonomic signature of celiac disease. *J Proteome Res* 2009;8:170-7.
30. Sellitto M, Bai G, Serena G, et al. Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants. *PLoS One* 2012;7:e33387.
31. Olivares M, Walker AW, Capilla A, et al. Gut microbiota trajectory in early life may predict development of celiac disease. *Microbiome* 2018;6:36.
32. Ilonen J, Kiviniemi M, Lempainen J, et al. Genetic susceptibility to type 1 diabetes in childhood—estimation of HLA class II associated disease risk and class II effect in various phases of islet autoimmunity. *Pediatr Diabetes* 2016;17:8-16.
33. Erkkola M, Karppinen M, Javanainen J, et al. Validity and reproducibility of a food frequency questionnaire for pregnant Finnish women. *Am J Epidemiol* 2001;154:466-476.
34. Virtanen SM, Nevalainen J, Kronberg-Kippilä C, et al. Food consumption and advanced β cell autoimmunity in young children with HLA-conferred susceptibility to type 1 diabetes: a nested case-control design-. *Am J Clin Nutr* 2012;95:471-478.
35. 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.
36. 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.
37. Carey RN, Wold S, Westgard JO. Principal component analysis: an alternative to "referee" methods in method comparison studies. *Anal Chem* 1975;47:1824-9.

38. Rosipal R, Krämer N. Overview and recent advances in partial least squares, In International Statistical and Optimization Perspectives Workshop" Subspace, Latent Structure and Feature Selection", Springer, 2005.
39. Tenenhaus M. La régression PLS: théorie et pratique: Editions technip, 1998.
40. Wold S, Sjöström M, Eriksson L. Partial least squares projections to latent structures (PLS) in chemistry. Encyclopedia of computational chemistry 2002;3.
41. Farrés M, Platikanov S, Tsakovski S, et al. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J Chemom* 2015;29:528-536.
42. Westerhuis JA, Hoefsloot HC, Smit S, et al. Assessment of PLSDA cross validation. *Metabolomics* 2008;4:81-89.
43. Pinheiro J, Bates D, DebRoy S, et al. Linear and nonlinear mixed effects models. R package version 3 2014.
44. Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS Springer. New York 2000.
45. Heiberger RM, Freeny AE, Chambers JM. Analysis of Variance; Designed Experiments. Statistical Models in S: Routledge, 2017:145-193.
46. Esposito C, Paparo F, Caputo I, et al. Anti-tissue transglutaminase antibodies from coeliac patients inhibit transglutaminase activity both in vitro and in situ. *Gut* 2002;51:177-181.
47. Hill P, McMillan S. Anti-tissue transglutaminase antibodies and their role in the investigation of coeliac disease. *Ann Clin Biochem* 2006;43:105-117.
48. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* 2018;46:D608-D617.
49. Agardh D, Lynch K, Brundin C, et al. Reduction of tissue transglutaminase autoantibody levels by gluten-free diet is associated with changes in subsets of peripheral blood lymphocytes in children with newly diagnosed coeliac disease. *Clin Exp Immunol* 2006;144:67-75.
50. Brar P, Kwon GY, Holleran S, et al. Change in lipid profile in celiac disease: beneficial effect of gluten-free diet. *Am. J. Med* 2006;119:786-790.
51. Kirchberg FF, Werkstetter KJ, Uhl O, et al. Investigating the early metabolic fingerprint of celiac disease—a prospective approach. *J Autoimmun* 2016;72:95-101.
52. Capristo E, Addolorato G, Mingrone G, et al. Low-serum high-density lipoprotein-cholesterol concentration as a sign of celiac disease. *Am. J. Gastroenterol* 2000;95:3331.
53. Farnetti S, Zocco MA, Garcovich M, et al. Functional and metabolic disorders in celiac disease: new implications for nutritional treatment. *J Med Food* 2014;17:1159-1164.
54. Ameer F, Scandiuzzi L, Hasnain S, et al. De novo lipogenesis in health and disease. *Metabolism* 2014;63:895-902.
55. Orešić M, Simell S, Sysi-Aho 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* 2008;205:2975-2984.
56. Orešić M, Gopalacharyulu P, Mykkänen J, et al. Cord serum lipidome in prediction of islet autoimmunity and type 1 diabetes. *Diabetes* 2013;DB_130159.

57. La Torre D, Seppänen-Laakso T, Larsson HE, et al. Decreased cord-blood phospholipids in young age at onset type 1 diabetes. *Diabetes* 2013;DB_130215.
58. Orešić M, Hyötyläinen T, Kotronen A, et al. Prediction of non-alcoholic fatty-liver disease and liver fat content by serum molecular lipids. *Diabetologia* 2013;56:2266-2274.
59. Luukkainen PK, Zhou Y, Sädevirta S, et al. Hepatic ceramides dissociate steatosis and insulin resistance in patients with non-alcoholic fatty liver disease. *J Hepatol* 2016;64:1167-1175.
60. Reilly NR, Lebwohl B, Hultcrantz R, et al. Increased risk of non-alcoholic fatty liver disease after diagnosis of celiac disease. *J Hepatol* 2015;62:1405-1411.
61. Haug K, Salek RM, Conesa P, et al. MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res* 2013;41:D781-6.

Table 1. Demographic and clinical characteristics of the study subjects.

	Celiac Disease	Healthy Controls
Number of subjects	23	23
Gender		
Male	12	12
Female	11	11
HLA DR-DQ genotype		
(DR ₃)-DQA1*05-DQB1*02/DRB1*04-DQA1*03-DQB1*0302	17	17
DRB1*04-DQA1*03-DQB1*0302/x ≠ DQB1*02, 03:01, or 06:02	6	6
Duration of exclusive breastfeeding (median)	1.4 months	1.6 months
Duration of breastfeeding (median)	8.0 months	8.0 months
Age at first gluten intake (median)	6.0 months	5.5 months
Age at first tTGA positivity (median) (seroconversion)	58 months	--
First tTGA level in plasma (median EIU)	116.8	--
Age at endoscopy (median)	58.5 months	--
tTGA level after GFD (median EIU)	7.1	--
Total follow-up age from birth (median)	6.1 years	6.2 years
Plasma samples analyzed by lipidomics	121	112

Figure captions

Figure 1. (a) PLS-DA score plot showing difference in the lipidome between CD progressors (cyan) and healthy controls (magenta). (b) A regression plot of the lipids and their VIP scores obtained from PLS-DA model. The lipids are grouped and color coded by their classes. (c) PLS-DA score plot showing difference in the lipidome in CD progressors along the age. (d) A regression plot of the lipids and their VIP scores at 3 months of age. The lipids are grouped and color coded by their classes.

Figure 2. (a) Heatmap showing \log_2 fold changes (FC) of significantly altered (FDR adjusted p-values < 0.05) longitudinal lipid profiles between CD progressors and matched healthy controls, as identified by linear mixed-effect (LME) models. Here, blue and red color depicts down- and up-regulated lipid intensities in CD progressor as compared to their matched healthy controls respectively, and white depicts no change. The lipids that changed specifically (HSD, p-values < 0.05) at a particular age, as identified by Tukey's test for HSD, is highlighted with a star **. (b-c) A correlation plot of log ration or fold changes of triacylglycerols (TGs) and the number of acyl chain bonds incurred by them at early (3 months) and later age (72 months), i.e., when gluten free diet has been started. R denotes ranked correlation coefficients. (d-e) Correlation plots of log ration of TGs and the number of acyl chain carbon atoms in TGs at the same age. The significantly changed ($p < 0.05$) TGs between CD progressors and healthy control (at any age) are marked as dark pink color.

Figure 3. Longitudinal profiles of lipids in CD progressors (black) and healthy controls (orange) with 95% confidence interval as shown by the area shaded around the curves.

Figure 4. (a) Heatmap showing the Spearman's rank coefficient estimated between significantly changed TG levels and tissue transglutaminase antibodies (tTGA) titer in the plasma of CD progressors immediately after the seroconversion (SC), and 6-12 months after the SC or introduction of gluten free diet (GFD). Red and blue color depicts positive and negative correlations. (b) Heatmap showing the Spearman's rank coefficient estimated

between significantly changed PC and plasma tTGA titer in CD progressors after SC and GFD.

Figure 5. (a-d) Boxplot showing total log intensities of essential TGs, nonessential TGs, PCs and SMs in the CD progressors and healthy controls before (3 months) and after introduction of gluten (12-36 months) in the diet and after introduction of GFD (72 months). (e) Gridmap of p-values obtained from ANOVA models by combining all possible conditions before/after the gluten intake in CD progressors and healthy controls are given. The three different conditions are, 'BA-Glut' = before and after gluten, 'AGlut-AGFD' = after gluten and after gluten free diet, 'BGlut-AGFD' = before gluten and after gluten free diet.

FIGURE 1

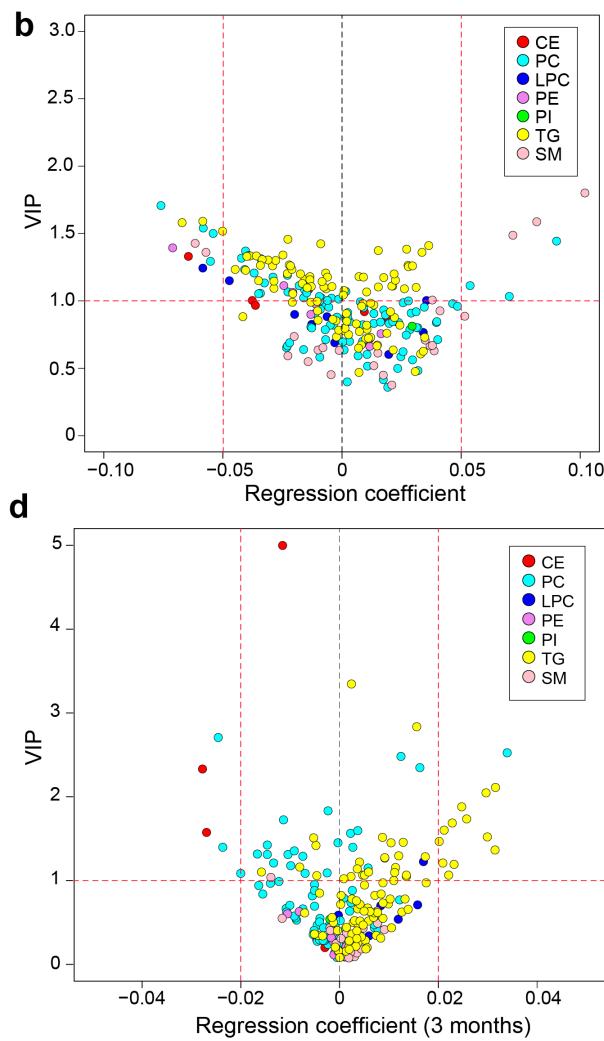
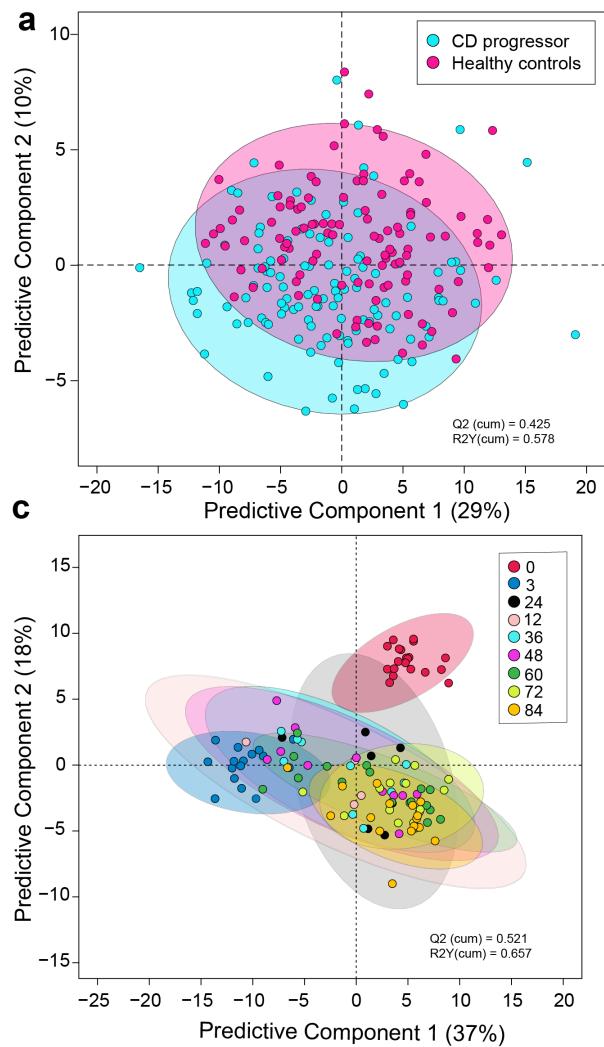


FIGURE 2

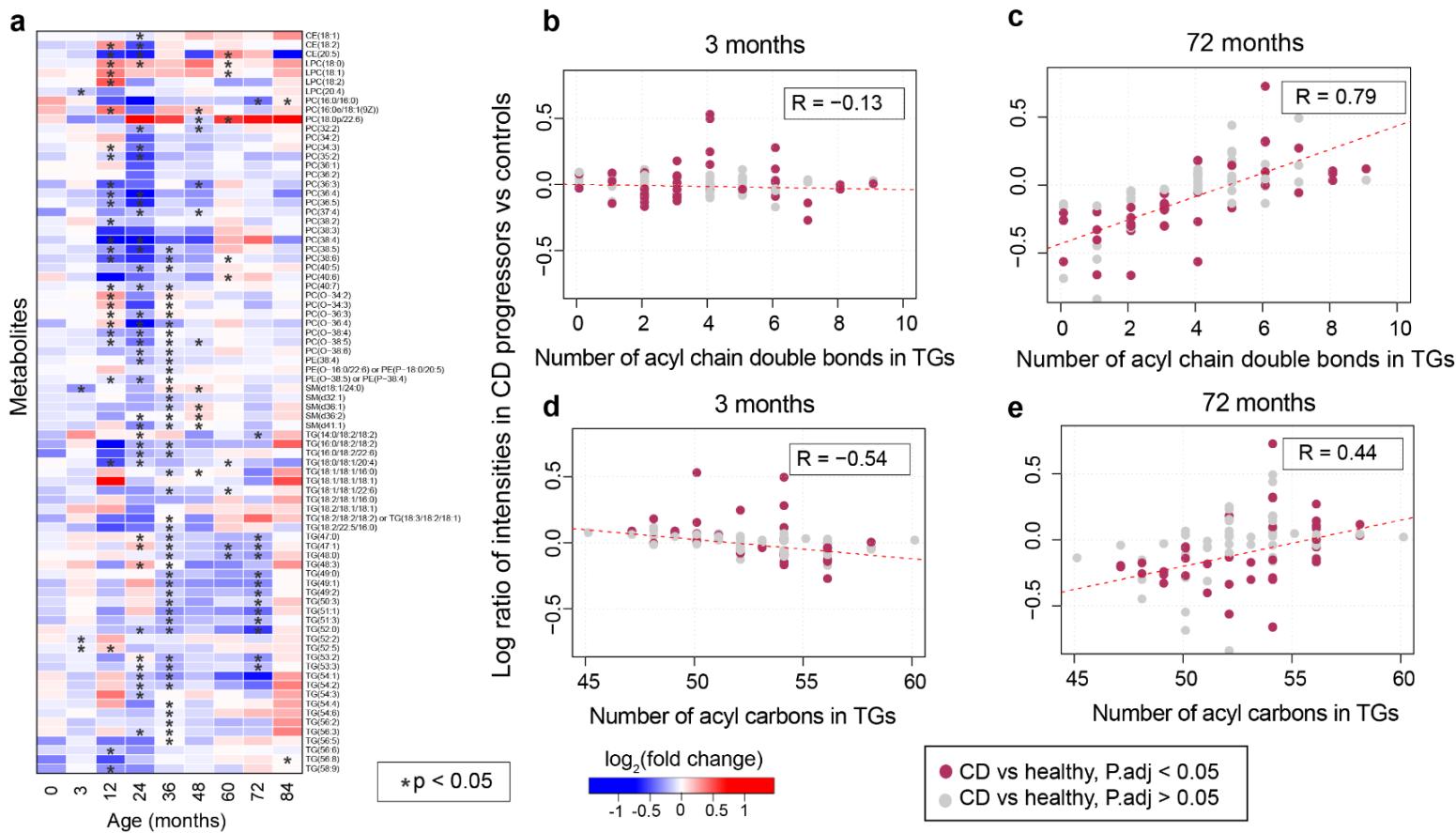


FIGURE 3

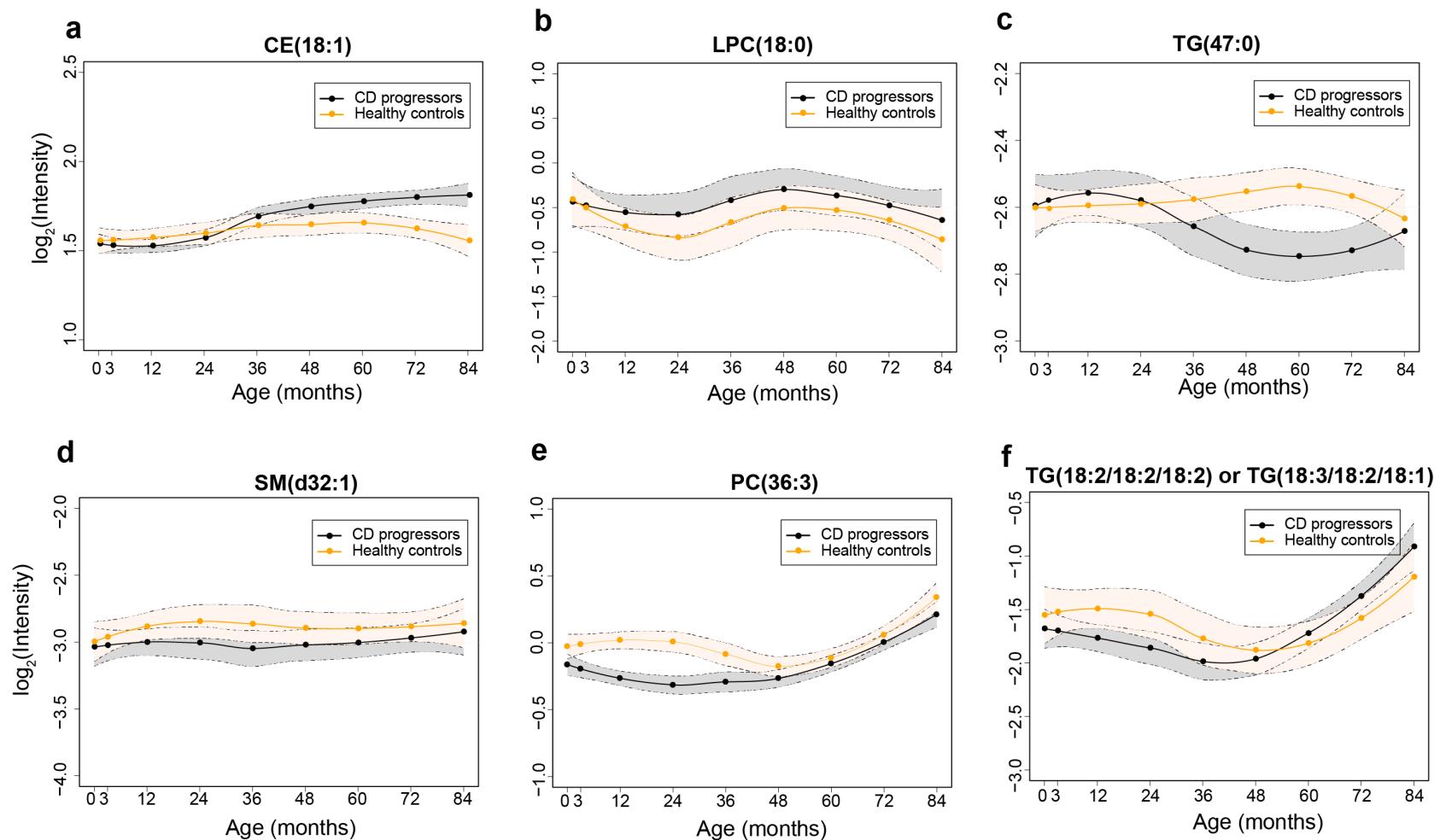


FIGURE 4

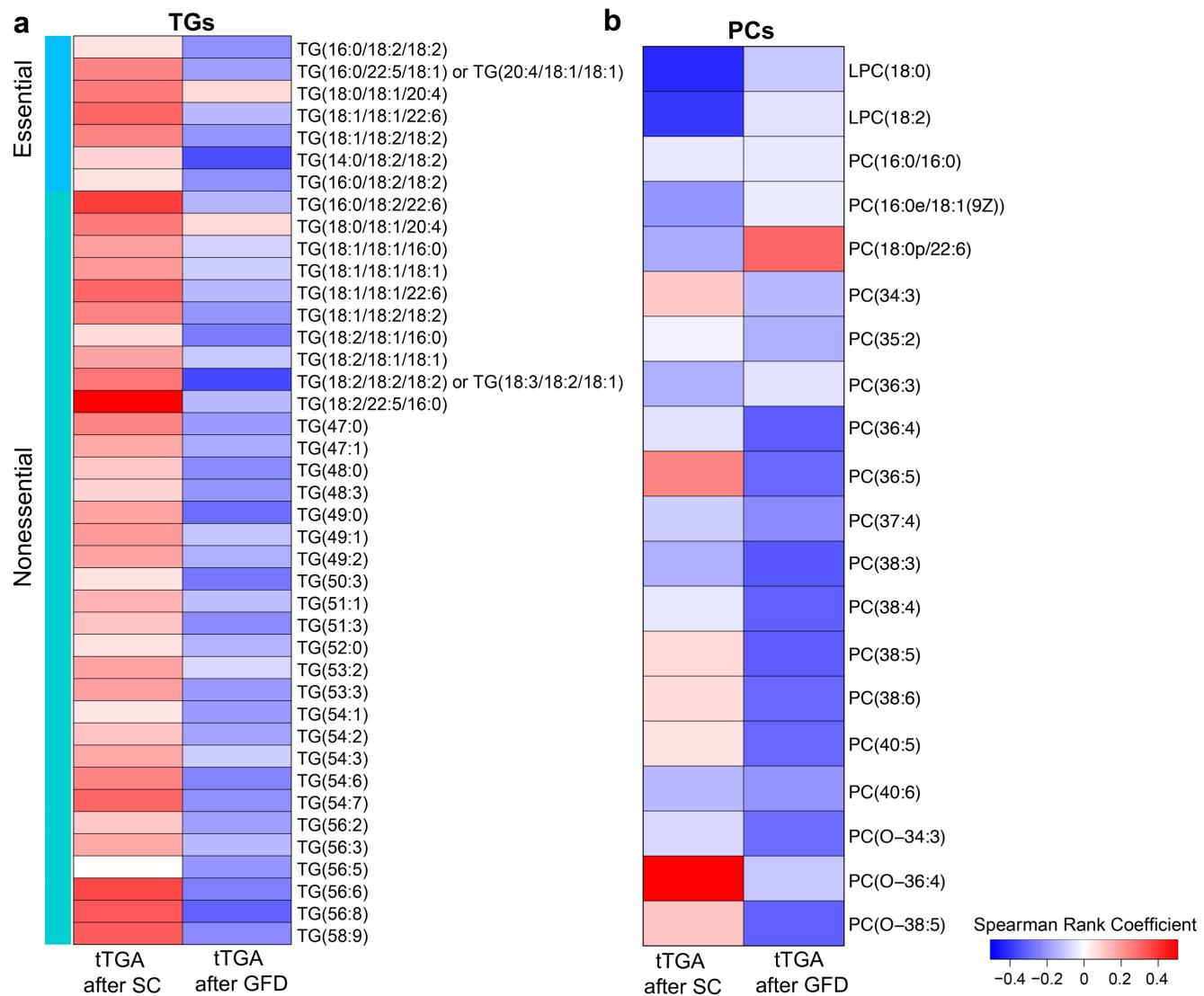


FIGURE 5

