Investigating the early metabolic fingerprint of Celiac Disease – a prospective approach.

Franca F. Kirchberg[1], Katharina J. Werkstetter[1], Olaf Uhl[1], Renata Auricchio[2], Gemma Castillejo[3], Ilma R. Korponay-Szabo[4], Isabel Polanco[5], Carmen Ribes Koninckx[6], Sabine L. Vriezinga[7], Berthold Koletzko[1,*], M. Luisa Mearin[7], Christian Hellmuth[1]

[1] Ludwig-Maximilians-Universität München, Division of Metabolic and Nutritional Medicine, Dr. von Hauner Children's Hospital, University of Munich Medical Center, Munich, Germany, [2]Dept. of Medical Translational Sciences and European Laboratory for the Investigation of Food-Induced Diseases, University Federico II, Naples, Italy, [3]Dept. of Pediatric Gastroenterology Unit, Hospital Universitari Sant Joan de Reus, URV, IIPV, Reus, Spain ,[4]Celiac Disease Center, Heim Pál Children's Hospital, Budapest, Hungary, [5]Dept. of Pediatric Gastroenterology and Nutrition, La Paz University Hospital, Madrid, [6]U. Enfermedad Celiaca e Inmunopatología Digestiva, Instituto de Investigación Sanitaria La Fe, Valencia, Spain, [7]Dept. of Pediatrics, Leiden University Medical Center, Leiden, The Netherlands,

Abbreviated title: Metabolic fingerprint of Celiac Disease

Role of the funding source:

This work was financially supported by the Commission of the European Communities, the 7th Framework Programme, contract FP7-289346-EARLY NUTRITION, and the European Research Council Advanced Grant ERC-2012-AdG – no.322605 META-GROWTH. This manuscript does not necessarily reflect the views of the Commission and in no way anticipates the future policy in this area. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare to have no conflict of interest.

Highlights:

- Serum of 4 month old infants is not predictive for development of coeliac disease up to school age
- Metabolic profiles are similar across the different HLA-risk groups
- Early diagnostic approaches should focus on children after the introduction of gluten

Corresponding author and person to whom reprint requests should be addressed:

Berthold Koletzko, Prof. of Paediatrics

Ludwig-Maximilians-University of Munich

Dr. von Hauner Children's Hospital, Univ. of Munich Medical Center

Lindwurmstr. 4, D-80337 Muenchen, Germany

Tel: +49 89 44005 2826, Fax: +49 89 44005 7742

E-mail: office.koletzko@med.lmu.de

Disclosure statement: The authors have nothing to disclose

Clinical Trial Registration Number: PreventCD Current Controlled Trials number, ISRCTN74582487.

Abstract

Objectives and study: In the development of Celiac Disease (CD) both genetic and environmental factors play a crucial role. The Human Leukocyte Antigen *(HLA)*-DQ2 and *HLA*-DQ8 loci are strongly related to the disease and are necessary but not sufficient for the development of CD. Therefore, increasing interest lies in examining the mechanisms of CD onset from the early beginning. Differences in serum and urine metabolic profiles between healthy individuals and CD patients have been reported previously. We aimed to investigate if the metabolic pathways were already altered in young, 4 month old, infants, preceding the CD diagnosis.

Methods: Serum samples were available for 230 four months old infants of the PreventCD project, a multicenter, randomized, double-blind, dietary intervention study. All children were positive for *HLA*-DQ2 and/or *HLA*-DQ8 and had at least one first-degree relative diagnosed with CD. Amino acids were quantified after derivatization with liquid chromatography – tandem mass spectrometry (MS/MS) and polar lipid concentrations (lyso-phosphatidylcholines, phosphatidylcholines, and sphingomyelins) were determined with direct infusion MS/MS.

We investigated the association of the metabolic profile with (1) the development of CD up to the age of 8 years (yes/no), (2) with HLA-risk groups, (3) with the age at CD diagnosis, using linear mixed models and cox proportional hazards models. Gender, intervention group, and age at blood withdrawal were included as potential confounder.

Results: By the end of 2014, thirty-three out of the 230 children (14%) were diagnosed with CD according to the ESPGHAN criteria. Median age at diagnosis was 3.4 years (IQR, 2.4 - 5.2). Testing each metabolite for a difference in the mean between healthy and CD children (1), we could not identify a discriminant analyte or a pattern pointing towards an altered metabolism (Bonferroni corrected $P > 0.05$ for all). Metabolite concentrations (2) did not differ across the HLA-risk groups. When including the age at diagnosis using (3) survival models, we found no evidence for an association between the metabolic profile and the risk of a later CD diagnosis.

Conclusion: The metabolic profile at 4 month of age was not predictive for the development of CD up to the age of 8 years. Our results suggest that metabolic pathways reflected in

serum are affected only later in life and that the HLA-genotype does not influence the serum metabolic profile in young infants before introduction of solid food.

Keywords: Celiac Disease, Human Leukocyte Antigen, Metabolomics, Early Programming

Introduction:

Celiac disease (CD) is a systemic immune-mediated disorder which is triggered by gluten and other prolamines in wheat, barley, rye, and other cereals in genetically susceptible individuals. CD has systemic effects but is mainly characterized by enteropathy with villous atrophy that may result in malnutrition, diarrhea, weight loss, growth retardation, and other symptoms such as constipation or fatigue [1]. The only known effective treatment is a lifelong, strict gluten-free diet (GFD) [2]. Environmental and genetic factors play an important role in the pathogenesis of CD. Apart from the main environmental trigger gluten, a protein complex formed by gliadins and glutenins [3], other environmental factors such as breastfeeding, time of gluten introduction, or the mode of delivery have been proposed to be associated with the risk for CD. While the influence of some of these environmental factors is being clarified or is still a question of debate [4, 5], the relevance of the risk alleles of the Human Leukocyte Antigen (HLA)-DQ2 and HLA-DQ8 for the development of CD is well established: Over 90% of CD patients carry the alleles encoding HLA-DQ2 molecule; most of the remaining CD patients carry the HLA-DQ8 heterodimers. However, although around 25% of the European general population is positive for the HLA-DQ2 heterodimer (and ~50% if HLA-DQ8 is also included), only a small fraction (~1%) will develop CD [6-8]. This indicates that the HLA-DQ2 and/or HLA-DQ8 risk alleles are necessary but not sufficient for the development of the disease [3]. Thus, other genetic and/or environmental factors besides gluten must be involved in the pathogenesis of CD and a closer investigation of the mechanisms involved in the activation and development of the disorder is needed.

Metabolomics, the study of small-molecule metabolite profiles, facilitates the characterization of several pathological conditions such as obesity or cardiovascular diseases [9, 10]. The metabolites are intermediates and end products of cellular regulatory processes, and their levels can be regarded to be the result of the interaction of genome, transcriptome, proteome, and the environment [11, 12]. Investigating the metabolic profile of CD patients is thus a logical consequence. So far, only few studies investigated the differences in metabolic profiles between CD patients and healthy controls [13], and even fewer studies focused on the serum metabolic profile. The results of those studies, however,

are quite promising revealing alterations in energy metabolism [14] and suggesting that metabolic alterations may precede the development of small intestinal villous atrophy [15].

Besides this cross-sectional view, it is also of major interest to investigate whether the metabolic profiles of children, who will progress to CD later in life, differ already at early age. The identification of metabolic markers would represent a significant advance and targets for early interventions and preventive strategies. For instance, studies on type 1 diabetes which shares common alleles with CD have identified metabolic phenotypes that characterize the early pathogenesis of the disease [16, 17]. Using data from the prospective cohort of the PreventCD study, we tested if the metabolic profile of 4 month old children at genetic risk for CD was associated with (1) the development of CD up to school age, (2) the HLA-risk groups, or (3) the age at CD diagnosis.

Material and Methods

Study design

The PreventCD project is a prospective, randomized, double-blind, placebo-controlled, dietary-intervention study in children with high risk for CD [5, 18]. The first child was included on May 26, 2007, and the follow-up for this analysis closed on August 26th, 2015. Recruitment of the infants was done consecutively through CD organizations in Croatia, Germany, Hungary, Israel, Italy, the Netherlands, Poland, and Spain. Newborns between 0–3 months of age were recruited if (i) they had at least one first-degree family member with CD confirmed by small-bowel biopsy, and (ii) were HLA-DQ2 or HLADQ8 positive, or otherwise carrying the allele DQB1*02 [18]. Premature infants or infants with syndromes associated with an increased risk of CD, such as trisomy 21 or Turner's syndrome, were excluded. The infants were randomized to the intervention groups and either received gluten (200 mg of vital wheat gluten mixed with 1.8 g of lactose) or placebo (2g of lactose) between 4 and 6 months of age. Randomization took place after blood withdrawal.

The study was approved by the medical ethics committee at each participating center and complied with Good Clinical Practice guidelines. The authors vouch for the veracity and completeness of the data and analyses reported and for the adherence of the study to the protocol.

Assessment of CD

Children with elevated levels of antibodies indicating CD and/or with clinical suspicion of CD are offered to undergo a small-bowel biopsy to diagnose the disorder. The diagnosis of CD was based on the histologic findings of the small-bowel biopsies, according to the criteria of the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) [19]. The age of the patient at the day of biopsy is considered to be the age at the diagnosis of CD.

Genotyping

Genotyping for HLA-DQ-alleles was performed by single nucleotide polymorphisms (SNPs) based on the tag-SNP approach (Department of Genetics University Medical Center Groningen, the Netherlands). The HLA-risk groups were defined as follows: (1) DR3-

DQ2/DR3-DQ2 (DQ2.5/DQ2.5); DR3-DQ2/DR7-DQ2 (DQ2.5/DQ2.2); (2) DR7- DQ2/DR5-DQ7 DQ2.2/DQ7); (3) DR3-DQ2/DR5-DQ7 (DQ2.5/DQ7); DR3-DQ2/DR4-DQ8 (DQ2.5/DQ8); DR3-DQ2/other (DQ2.5/other); (4) DR7-DQ2/DR7-DQ2 (DQ2.2/DQ2.2); DR7-DQ2/DR4-DQ8 (DQ2.2/DQ8); DR4- DQ8/DR4-DQ8; (DQ8/DQ8)(5) DR7-DQ2/other (DQ2.2/other); DR4-DQ8/DR5-DQ7 (DQ8/DQ7); DR4-DQ8/other (DQ8/other), where other refers to any HLA-DQ haplotype except for DR3-DQ2, DR7-DQ2, DR4-DQ8 or DR5-DQ7 [5, 20]:

Sample collection & quantification of metabolites

Measurements of serum antigliadin antibodies, TG2A, and metabolomics were performed centrally. Blood samples were collected and centrifuged and the frozen serum samples (-20°C) were sent for determination of CD antibodies to Thermo Fisher Scientific, Freiburg, Germany. The rest of the serum was stored at -20°C and sent to the central sera bank of the project at the department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, where it was stored also at-20°C. Serum samples were available for 230 children who previously have been selected for a study on breastmilk composition during the first months of life and serum composition at 4 month (paper under submission). We took advantage of this collection for the purpose of this study. The sera of these 4 month old infants collected before the start of the dietary intervention were transferred on dry ice to LMU Munich and stored at -80°C until metabolomic analysis.

Analysis of AA was performed as described previously [21]. Briefly, 10µl of serum was prepared by derivatization to amino acid butylester, and determined by ion-pair liquid chromatography (LC). The LC system (Agilent, Waldbronn, Germany) was coupled to a tandem mass spectrometer (MS/MS, QTRAP4000, Sciex, Darmstadt, Germany) with an atmospheric pressure chemical ionization source operating in positive ionization mode. The MS was run in Multiple Reaction Monitoring (MRM) mode.
For quantification of polar lipids, namely free carnitine, acylcarnitines (Carn), lysophosphatidylcholines (LPC), diacyl-phosphatidylcholines (PCaa), acyl-alkyl-phosphatidylcholines (PCae), and sphingomyelines (SM), serum was diluted with methanol and, after centrifugation, supernatants were injected into the LC-MS/MS system with an electrospray ionization source. The system was used as flow-injection analysis and the MS

was run in MRM mode in both positive and negative ionization mode. The structures of the hydrocarbon chains of the polar lipids are indicated by the denotation Cx:y where x and y denote the length of the hydrocarbon chain and the number of unsaturations in the chain, respectively. 'a' indicates that the chain is bound via an ester bond to the backbone, while 'e' indicates binding by an ether bound. 'OH' stands for an additional hydroxyl group and 'DC' for two carboxyl groups in the molecule.

During the quantification processes, the samples of the study centers were randomly distributed across the batches. The entire analytical process was post-processed by the softwares Analyst 1.5.1 and R 3.2.1. To remove the batch effect, we centered the metabolomics data for analysis by subtracting for each metabolite its mean concentration across all batches. Analytes, for which we had measurements for less than 10% of the infants only, were excluded from the analyses. All metabolite concentrations are reported in µmol/L.

Statistical analyses

In a first step, the data was examined graphically and screened for outliers. We defined a measurement as outlier if its absolute concentration exceeds one standard deviation (SD) away from its neighbor. Baseline characteristics of the children are presented as mean (SD) or count (percentage), for continuous and categorical variables, respectively.

In a first analysis, we investigated if the metabolic profile was associated with the development of CD up to the age of 8 years, coded binary (yes/no). In a second analysis, we examined if the metabolic profile depended on the HLA-risk group. In both analyses, we used linear mixed models (LMM) with random intercept for study center and regressed the metabolite's concentration on the (1) binary CD y/n variable and the (2) categorical HLA-risk group variable. When analyzing the sum parameters, we used the non-batch centered data and calculated LMM with two independent random intercepts for study center and batch number. As the HLA-risk groups sizes were very unequal, we chose the biggest, namely risk group 3, as reference category. ANOVA was used to test for differences between the HLA-risk groups. Possible confounding factors that were considered in both analyses comprised gender, maternal CD status, and age at blood withdrawal. We furthermore applied cox proportional hazards models where we investigated if the metabolite concentration was

associated with the time to the CD diagnosis (i.e. the age at diagnosis). In these models, we additionally considered intervention group as confounding factor. Bonferroni correction was applied to all analyses in order to control for false positives. Thus, a *P*-value of below 0.05/183 = 0.00027 was considered to be statistically significant. Furthermore, as the aforementioned LMM only consider one metabolite at a time, we performed sensitivity analyses using partial least squares discriminant analysis and random forest. Random forest was calculated the same way as described in Kirchberg et al. [22].

The three analyses were each performed on the whole sample and on a subgroup, defined as follows: Firstly, the infant had to be exclusively breastfed up to the age of 4 month; Secondly, the mother had to be either on a GFD and being CD positive, or not being on a GFD and being healthy (n=174).

We presented our results graphically in a circular Manhattan Plot summarizing the results of the three approaches. Every metabolite is given a position on the outer circle and three points, each indicating a $-\log_{10}(P)$ value from a model, are positioned on the line connecting the metabolite to the center. Significance can be read by looking at the significance lines at $\alpha=0.05$ and $\alpha=0.00027$ (Bonferroni corrected). If a point lies further out than the line, it is significant to the corresponding significance level α.

Results

Blood samples were available for 230 children. Baseline characteristics of those infants are listed in Table 1. The mean age at blood withdrawal was 122.3 days (SD, 25.72 d). Up to that day, 184 children (80%) were exclusively breastfed. Among those, 90 mothers were diagnosed with CD and were on a GFD diet, and 84 mothers were healthy and hence not on a GFD diet. The remaining 10 mothers were either on a GFD although being healthy (n=1), were not on a GFD although having been diagnosed with CD (n=4), or information on their diet was missing (n=5). By the end of 2014, and when the children were aged on average 6.5 years (SD, 0.69 yrs), thirty-three out of these 230 children (14%) were diagnosed with CD according to the ESPGHAN criteria. The mean age at diagnosis was 3.8 years (median, 3.4 yrs; IQR, 2.3-5.2 yrs) and youngest age at diagnosis was 1.8 years of age. Of the 174 children who were exclusively breastfed and whose mother subsisted according to her CD status, 23 (13%) were diagnosed with CD. Children comprised in the subsample were not different to those not comprised with respect to the baseline characteristics listed in Table 1 (data not shown). Regarding the HLA-risk groups, 13% were in the highest risk group. Most children were in the risk group 3 (52%).

Metabolic profile in relation to a later CD diagnosis and the HLA-risk group

The metabolic profiles between children with and without a later CD diagnosis or between the five HLA-risk groups were not different (all corrected $P = 1$; Supplemental Table 1 & 2). This result did not change in the multivariate LMM where we additionally adjusted for gender, maternal CD status, and age at blood withdrawal. The *P* –values for both analyses are represented graphically in the circular Manhattan-Plot (Figure 1): Although some metabolites differed at the uncorrected significance level of $\alpha=0.05$, these differences were far away from significance after correcting for multiple testing. Table 2 and 3 summarize the concentrations of these metabolites. Regarding the differences between infants with and without development of CD up to school age (Table 2) all of these metabolites are phosphatidylcholines. Among children who were exclusively breastfed up to 4 month of age and whose mothers had a diet consistent to the CD status (thus, were on a GFD and had a CD diagnosis, or were not on a GFD but had a CD diagnosis), this observation becomes even clearer: Some metabolites almost reached the Bonferroni-corrected significance level

(Supplemental Figure 1). For instance, the corrected P-value comparing the concentration of the acyl-alkyl phosphatidylcholine PCae C42:6 between children who will and those who won't develop CD in comparison to healthy infants was 0.08. The calculation of the absolute sum of the phosphatidylcholine species (Figure 2) revealed that the mean sums were lower in those infants who will develop CD later in life. This difference, however, was not significant at the corrected significance level, neither in the whole nor in the subsample (Supplemental Figure 2). The (percent) phospholipid composition also was not different (Supplemental Table 3).

The ensemble approach, the random forest analysis, also confirmed the non-discriminating nature of the metabolome and dispels the impression from the subsample analysis, this is, different phospholipid levels in healthy and CD children: The out-of-bag error rate of 13% (11% in the subsample) equaled in both cases the percentage of CD diagnosis among the studied infants. Thus, random forest was not able to classify correctly CD and healthy children but classified everyone as healthy.

Table 3 summarizes the results of selected LMM on the HLA-risk groups. The concentrations are similar across the groups (all corrected P-values > above 0.05).

Relating the metabolic profile at 4 month of age with the age at diagnosis

Using cox proportional hazards models, we modelled the predictive power of the metabolite's concentrations on the time to onset of CD. No significant effect was found after Bonferroni correction (Figure 1). In fact, significances were similar to those of the LMM which ignores missing data originating from the censoring. Table 2 summarizes the hazard ratios and significances of selected metabolites and underlines this similarity.

Discussion

In this investigation on the metabolic preconditions of CD, neither a single metabolite nor any tested combination of metabolites was associated with the CD status at school age, or with the age at CD diagnosis. The HLA-risk group also had no effect on the metabolites' serum concentration.

To our knowledge, there have been no previous studies that examined the associations of early metabolic markers in the blood with a later CD diagnosis. Some publications focus on the metabolic fingerprint for CD in one of the matrices serum, urine, or gut microbiota, but they compare CD patients (potential or overt) with healthy controls [14, 15]. Their design differs from ours in that it is cross sectional rather than prospective. Calabro et al [13] provide a summary of these studies and conclude that there is clear evidence for an altered metabolism with respect to malabsorption, energy metabolism, and alterations in gut microflora and/or intestinal permeability. Two studies investigated the serum metabolic profile: Bertini [14] found differences in the plasma metabolic profiles of CD patients and healthy controls; particularly glycolysis was reduced in untreated CD patients. The consequently increased beta-oxidation along with malabsorption was suggested to be the reason for decreased lipids in sera and an increase in the ketone body 3-hydroxybutyric acid. Introducing a GFD led to a normalization of the main energy metabolic pathway and a recovery of villous functioning. Bernini et al [15] put emphasis on the precondition of CD and compared healthy subjects with those having potential or overt CD. Potential CD patients, having the typical immunological features of clear CD in the absence of any apparent intestinal damage, were found to have a virtually similar metabolic profile as the patients with overt CD. They suggest that metabolic and enterocyte alterations of CD arise before the observable modification of jejunal mucosa.

In contrast to these studies, we focused on the alteration of pathways prior to the presence of specific CD antibodies (IgA tissue transglutaminase) or CD onset and, according to the best of our knowledge, we are the first to examine potential metabolic preconditions for CD development in very young infants at genetic risk. Although the results of the cited previous studies point towards alterations in the metabolic profile of CD patients, our findings do not show any such changes in infants aged 4 months prior to dietary gluten introduction. We

observed a trend to decreased phospholipid levels in children who later developed CD. However, we strongly emphasize that all corrected *P*-values were above 0.05, no matter which analysis we performed, including those for the sums of the phospholipid species. As the results were stronger in the subsample of mainly exclusively breastfed children, it seems that the consumption of liquids or foods other than breastmilk reduces this trend. To exclude a potential influence of a different maternal diet and breastmilk composition of CD mothers, we adjusted the analyses for the maternal CD status and gluten-free diet, but this was also not associated to any metabolite concentration. Decreased phospholipid levels, including phosphatidylcholines and sphingomyelins which both contain fatty acid (FA) chains, have been found in cord blood of children who later progress to type 1 diabetes, a disease which is also associated with the HLA class II genes [17, 23]. There are no comparable studies on CD. Studies on the FA composition in CD patients have found ambiguous results: While some found evidence for an altered serum FA composition in adults with CD [24-26], Steel et al [27] found a similar serum profile of FA in phospholipids in CD and control children. We could not identify any distinct FA species among the phosphatidylcholines (Table 2). Phospholipid composition was also not associated to the CD status at school age. Tjellström et al [28] who focus on children who already developed CD found a different fecal pattern of some short chain fatty acids (SCFA) between untreated CD children and healthy controls which may reflect bacterial metabolism of undigested carbohydrates secondary to CD-induced malabsorption. This however is not expected to exists at the age of 4 month prior to gluten introduction. SCFA are products of microbial degradation of complex carbohydrates in the colon and, while butyrate mainly feeds the enterocytes, acetate and propionate are transported to peripheral organs for gluconeogenesis, lipogenesis and other metabolic and immune response pathways [29, 30]. It is very likely that this effect would have led to respective alterations in the patterns of esterified serum lipids and acylcarnitines which we did not observe. Therefore, we assume that no major alterations in SCFA are present in 4 month old infants with high risk for CD before gluten introduction. However, further studies measuring serum SCFA are needed to gain certainty on this point.

Regarding the genotype, there were no metabolomic differences across the HLA-risk groups, thus the HLA-DQ genotype does not directly cause the altered metabolism observed later in CD patients. The main genetic predisposing factors are the HLA-DQ alleles accounting for

18% of the genetic risk [31]. Individuals with HLA-DQ2/DQ-8 haplotypes have increased immunological responsiveness to prolamins such as dietary wheat gliadin due to the CD-associated HLA-DQ molecules that bind gluten peptides with a superior ability than the other HLA-molecules [3]. It was found that the HLA-DQ genotype influences the early gut microbial colonization [32, 33]. As the metabolome is influenced by the microbiome [34], these findings suggest an altered metabolic profile according to the HLA-risk group which however was not confirmed by this study of infants early before onset of CD. Besides the HLA-genes, also other loci are associated to CD. The total genetic variation is estimated to explain up to 48% of CD heritability [31]. Regulatory mechanisms of CD development may also be modulated by epigenetics and microbiota [6, 35].

The strength of this study lies in its prospective character, the long follow-up of up to 8 years, and in the broad assessment of CD related factors, such as information on the mother's diet and CD status as well as the child's HLA-genotype. CD cases were all assessed in an identical way which minimizes the risk of bias. The majority of the infants were exclusively breastfed up to blood withdrawal. Having a sufficient sample size to analyze the children with exclusive breastfeeding and hence to reduce diet induced metabolic variation, we were able to investigate the association of the metabolome not yet influenced by liquids or foods other than breastmilk and future CD development.

Conclusions

Our findings do not support a relation of the metabolic profile at age of 4 month prior to gluten-consumption, neither with respect to a future CD development nor with respect to the genotype. Performing metabolomics profiling longitudinally at several time points might help to better capture the dynamics of the CD development. With respect to the attempts to understand and prevent the development of CD, our results lead us to the conclusion that metabolic investigations should focus on children after the introduction of gluten.

Acknowledgements

We thank Stephanie Winterstetter, who analyzed the blood plasma samples; Yvonne Wijkhuisen, project manager of PreventCD, for support; Els Stoopman, who helped with the data management; and all the children and families who participated in this project. The data presented are part of the PhD thesis accomplished by Franca Kirchberg at the Medical Faculty of the Ludwig-Maximilians-University of Munich.

References

[1] Husby S, Koletzko S, Korponay-Szabo IR, Mearin ML, Phillips A, Shamir R *et al.* European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. Journal of pediatric gastroenterology and nutrition, 2012;54:136-60.

[2] Green PH, Jabri B. Coeliac disease. Lancet, 2003;362:383-91.

[3] Sollid LM, Lie BA. Celiac disease genetics: current concepts and practical applications. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association, 2005;3:843-51.

[4] Chmielewska A, Piescik-Lech M, Szajewska H, Shamir R. Primary Prevention of Celiac Disease: Environmental Factors with a Focus on Early Nutrition. Annals of nutrition & metabolism, 2015;67 Suppl 2:43-50.

[5] Vriezinga SL, Auricchio R, Bravi E, Castillejo G, Chmielewska A, Crespo Escobar P *et al.* Randomized feeding intervention in infants at high risk for celiac disease. The New England journal of medicine, 2014;371:1304-15.

[6] Dieli-Crimi R, Cenit MC, Nunez C. The genetics of celiac disease: A comprehensive review of clinical implications. Journal of autoimmunity, 2015;64:26-41.

[7] Johnson TC, Diamond B, Memeo L, Negulescu H, Hovhanissyan Z, Verkarre V *et al.* Relationship of HLA-DQ8 and severity of celiac disease: comparison of New York and Parisian cohorts. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association, 2004;2:888-94.

[8] Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. The Journal of experimental medicine, 1989;169:345-50.

[9] Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant molecular biology, 2002;48:155-71.

[10] Hivert MF, Perng W, Watkins SM, Newgard CS, Kenny LC, Kristal BS *et al.* Metabolomics in the developmental origins of obesity and its cardiometabolic consequences. Journal of developmental origins of health and disease, 2015;6:65-78.

[11] Krug S, Kastenmuller G, Stuckler F, Rist MJ, Skurk T, Sailer M *et al.* The dynamic range of the human metabolome revealed by challenges. FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2012;26:2607-19.

[12] Nicholson JK, Lindon JC. Systems biology: Metabonomics. Nature, 2008;455:1054-6.

[13] Calabro A, Gralka E, Luchinat C, Saccenti E, Tenori L. A metabolomic perspective on coeliac disease. Autoimmune diseases, 2014;2014:756138.

[14] Bertini I, Calabro A, De Carli V, Luchinat C, Nepi S, Porfirio B *et al.* The metabonomic signature of celiac disease. Journal of proteome research, 2009;8:170-7.

[15] Bernini P, Bertini I, Calabro A, la Marca G, Lami G, Luchinat C *et al.* Are patients with potential celiac disease really potential? The answer of metabonomics. Journal of proteome research, 2011;10:714-21.

[16] La Torre D, Seppanen-Laakso T, Larsson HE, Hyotylainen T, Ivarsson SA, Lernmark A *et al.* Decreased cord-blood phospholipids in young age-at-onset type 1 diabetes. Diabetes, 2013;62:3951-6.

[17] Oresic M, Simell S, Sysi-Aho M, Nanto-Salonen K, Seppanen-Laakso T, Parikka V *et al.* Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. The Journal of experimental medicine, 2008;205:2975-84.

[18] Hogen Esch CE, Rosen A, Auricchio R, Romanos J, Chmielewska A, Putter H *et al.* The PreventCD Study design: towards new strategies for the prevention of coeliac disease. European journal of gastroenterology & hepatology, 2010;22:1424-30.

[19] Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. Archives of disease in childhood, 1990;65:909-11.

[20] Bourgey M, Calcagno G, Tinto N, Gennarelli D, Margaritte-Jeannin P, Greco L *et al.* HLA related genetic risk for coeliac disease. Gut, 2007;56:1054-9.
[21] Harder U, Koletzko B, Peissner W. Quantification of 22 plasma amino acids combining derivatization and ion-pair LC-MS/MS. Journal of chromatography B, Analytical technologies in the biomedical and life sciences, 2011;879:495-504.
[22] Kirchberg FF, Harder U, Weber M, Grote V, Demmelmair H, Peissner W *et al.* Dietary protein intake affects amino acid and acylcarnitine metabolism in infants aged 6 months. The Journal of clinical endocrinology and metabolism, 2015;100:149-58.
[23] Oresic M, Gopalacharyulu P, Mykkanen J, Lietzen N, Makinen M, Nygren H *et al.* Cord serum lipidome in prediction of islet autoimmunity and type 1 diabetes. Diabetes, 2013;62:3268-74.
[24] Jaskiewicz J, Szafran H, Kruszewska M, Brylska U, Krol M. The effect of gluten-free diet supplemented with Humana-MCT on the level of lipid fractions in blood serum of infants with coeliac disease. Acta physiologica Polonica, 1987;38:22-30.
[25] Rey J, Frezal J, Polonovski J, Lamy M. [Modifications of plasma lipids in disorders of intestinal absorption in children]. Revue francaise d'etudes cliniques et biologiques, 1965;10:488-94.
[26] Solakivi T, Kaukinen K, Kunnas T, Lehtimaki T, Maki M, Nikkari ST. Serum fatty acid profile in celiac disease patients before and after a gluten-free diet. Scandinavian journal of gastroenterology, 2009;44:826-30.
[27] Steel DM, Ryd W, Ascher H, Strandvik B. Abnormal fatty acid pattern in intestinal mucosa of children with celiac disease is not reflected in serum phospholipids. Journal of pediatric gastroenterology and nutrition, 2006;43:318-23.
[28] Tjellstrom B, Stenhammar L, Hogberg L, Falth-Magnusson K, Magnusson KE, Midtvedt T *et al.* Gut microflora associated characteristics in children with celiac disease. The American journal of gastroenterology, 2005;100:2784-8.
[29] Kasubuchi M, Hasegawa S, Hiramatsu T, Ichimura A, Kimura I. Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation. Nutrients, 2015;7:2839-49.
[30] Tremaroli V, Backhed F. Functional interactions between the gut microbiota and host metabolism. Nature, 2012;489:242-9.
[31] Gutierrez-Achury J, Zhernakova A, Pulit SL, Trynka G, Hunt KA, Romanos J *et al.* Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. Nature genetics, 2015;47:577-8.
[32] Olivares M, Albrecht S, De Palma G, Ferrer MD, Castillejo G, Schols HA *et al.* Human milk composition differs in healthy mothers and mothers with celiac disease. European journal of nutrition, 2015;54:119-28.
[33] Palma GD, Capilla A, Nova E, Castillejo G, Varea V, Pozo T *et al.* Influence of milk-feeding type and genetic risk of developing coeliac disease on intestinal microbiota of infants: the PROFICEL study. PloS one, 2012;7:e30791.
[34] Goodacre R. Metabolomics of a superorganism. The Journal of nutrition, 2007;137:259S-66S.
[35] Nistico L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG *et al.* Concordance, disease progression, and heritability of coeliac disease in Italian twins. Gut, 2006;55:803-8.

Tables

Table 1. Characteristics of the 230 children with metabolite measurements at the age of 4 month. Values are given as n (%), if not indicated otherwise.

	N = 230
Later CD Diagnosis [*]	
No	197 (86%)
Yes	33 (14%)
Female	115 (50%)
Exclusively breastfed up to age 4 month	184 (80%)
Mean age in days at blood withdrawal, (SD)	122.3 (25.7)
HLA-risk group [†]	
1	30 (13%)
2	19 (8%)
3	118 (52%)
4	10 (4%)
5	50 (22%)
Mother with CD	
No	116 (50%)
Yes	114 (50%)
Mother on GFD	
No	117 (52%)
Yes	108 (48%)

Abbreviations: CD, celiac disease; GFD, gluten-free diet; HLA, Human Leukocyte Antigen; SD, standard deviation

[*] Children were aged 6.5 on average when follow-up was closed for this analysis.

[†]Data on the HLA-risk groups were available for 227 children, with HLA typing performed by means of single-nucleotide polymorphisms (SNPs) on the basis of the tag-SNP approach. For the remaining 3 children, the status with regard to HLA-DQ2 and HLA-DQ8 positivity was determined by means of the Eu-Gen Risk test (Eurospital), with no information provided regarding the HLA-risk group. HLA-risk groups were defined as follows: (1) DR3-DQ2/DR3-DQ2 (DQ2.5/DQ2.5); DR3-DQ2/DR7-DQ2 (DQ2.5/DQ2.2); (2) DR7- DQ2/DR5-DQ7 DQ2.2/DQ7); (3) DR3-DQ2/DR5-DQ7 (DQ2.5/DQ7); DR3-DQ2/DR4-DQ8 (DQ2.5/DQ8); DR3- DQ2/other (DQ2.5/other); (4) DR7-DQ2/DR7-DQ2 (DQ2.2/DQ2.2); DR7-DQ2/DR4-DQ8 (DQ2.2/DQ8); DR4- DQ8/DR4-DQ8; (DQ8/DQ8)(5) DR7-DQ2/other (DQ2.2/other); DR4-DQ8/DR5-DQ7 (DQ8/DQ7); DR4-DQ8/other (DQ8/other), where other refers to any HLA-DQ haplotype except for DR3-DQ2, DR7-DQ2, DR4-DQ8 or DR5-DQ7

Table 2: Mean (SD) of selected serum metabolites of 4 month old infants that did ('CD') or did not ('healthy') develop celiac disease (CD) later in life. *P*-values were calculated using linear mixed models (LMM).

	CD status		Results from LMM				Results from Cox Proportional Hazards Model					
			Unadjusted models		Adjusted models*		Unadjusted models			Adjusted models†		
Metabolites (µM)	Healthy (n=197)	CD (n=33)	*P*	Bonferroni-corrected *P*	*P*	Bonferroni-corrected *P*	HR	*P*	Bonferroni-corrected *P*	HR	*P*	Bonferroni-corrected *P*
Diacyl-Phosphatidylcholines												
PCaa C30:4	0.062 (0.02)	0.054 (0.02)	0.048	1	0.052	1	0.98‡	0.043	1	0.98‡	0.036	1
PCaa C40:4	4.6 (1.31)	4 (1.11)	0.011	1	0.008	1	0.71	0.013	1	0.69	0.008	1
PCaa C42:5	1 (0.3)	0.89 (0.3)	0.017	1	0.011	1	0.23	0.016	1	0.21	0.01	1
Acyl-Alkyl-Phosphatidylcholines												
PCae C36:0	1.3 (0.4)	1.1 (0.3)	0.0495	1	0.033	1	0.39	0.037	1	0.36	0.023	1
PCae C38:0	4.9 (1.7)	4.2 (1.59)	0.0496	1	0.024	1	0.77	0.028	1	0.75	0.018	1
PCae C38:3	18 (5.12)	16 (4.77)	0.037	1	0.036	1	0.93	0.04	1	0.93	0.034	1
PCae C40:1	3.8 (1.33)	3.3 (1.13)	0.038	1	0.029	1	0.73	0.035	1	0.72	0.024	1
PCae C40:5	9.1 (3.04)	8 (2.73)	0.039	1	0.017	1	0.87	0.032	1	0.86	0.02	1
PCae C40:6	3.5 (1.13)	3.1 (1.15)	0.045	1	0.025	1	0.69	0.023	1	0.67	0.017	1
PCae C42:0	2.2 (1.03)	1.8 (0.8)	0.047	1	0.045	1	0.68	0.055	1	0.67	0.046	1
PCae C42:3	2.9 (0.9)	2.4 (0.8)	0.014	1	0.009	1	0.59	0.011	1	0.58	0.007	1
PCae C42:6	1.8 (0.5)	1.6 (0.4)	0.015	1	0.008	1	0.4	0.011	1	0.37	0.006	1

Abbreviations: CD, celiac disease; HR, Hazard Ratio; PCaa, diacyl-phosphatidylcholines; PCae, acyl-alkyl-phosphatidylcholines; SD, standard deviation

* Adjusted for gender, maternal CD status, and age at blood withdrawal

† Adjusted for gender and intervention group

‡ Hazard ratio per nmol/L increase

Table 3: Mean (SD) of selected serum metabolites of 4 month old infants in the HLA-risk groups[1]. *P*-values to test for differences across the groups were calculated using linear mixed models (LMM).

Metabolites (µM)	HLA-risk groups*					*P*-values from LMM			
						Unadjusted models		Adjusted models†	
	1 (n=30)	2 (n=19)	3 (n=118)	4 (n=10)	5 (n=50)	*P*	Bonferroni-corrected *P*	*P*	Bonferroni-corrected *P*
Carn C8:1	0.094 (0.06)	0.096 (0.06)	0.11 (0.08)	0.16 (0.1)	0.13 (0.09)	0.049	1	0.045	1
LPC C18:3	0.28 (0.1)	0.28 (0.1)	0.3 (0.1)	0.41 (0.3)	0.31 (0.1)	0.042	1	0.06	1

Abbreviations: Carn, acylcarnitine; CD, celiac disease; HLA, Human Leukocyte Antigen; LPC, lysophosphatidylcholine; SD, standard deviation

* HLA typing was performed by means of single-nucleotide polymorphisms (SNPs) on the basis of the tag-SNP approach. HLA-risk groups were defined as in Vriezinga (2014).

† Adjusted for gender, maternal CD status, and age at blood withdrawal

Figure Legend

Figure 1: Circular Manhattan Plot picturing the $-\log_{10}$(*P*-values) of the following three models: (1) linear mixed model (LMM) testing for the difference in the respective serum metabolite between infants that did or did not develop celiac disease ("CD") later in life; (2) LMM testing for the differences in the respective serum metabolite across the HLA-risk groups as defined in Bourgey (2007) ("HLA"); (3) Cox proportional hazards models assessing the effect of the respective metabolite on age at CD diagnosis ("Cox"). LMM were adjusted for gender, maternal CD status, and age at blood withdrawal. Cox proportional hazards models were adjusted for gender and intervention group. Significance is indicated by the grey dashed line (α=0.05), and the black dashed line (Bonferroni-corrected α=0.000027).The center represents an uncorrected *P*-value of 1.

Abbreviations: Carn, acylcarnitine; LPC, lysophosphatidylcholine; PCaa, diacyl-phosphatidylcholine; PCae, acyl-alkyl-phosphatidylcholine; SM, sphingomyeline

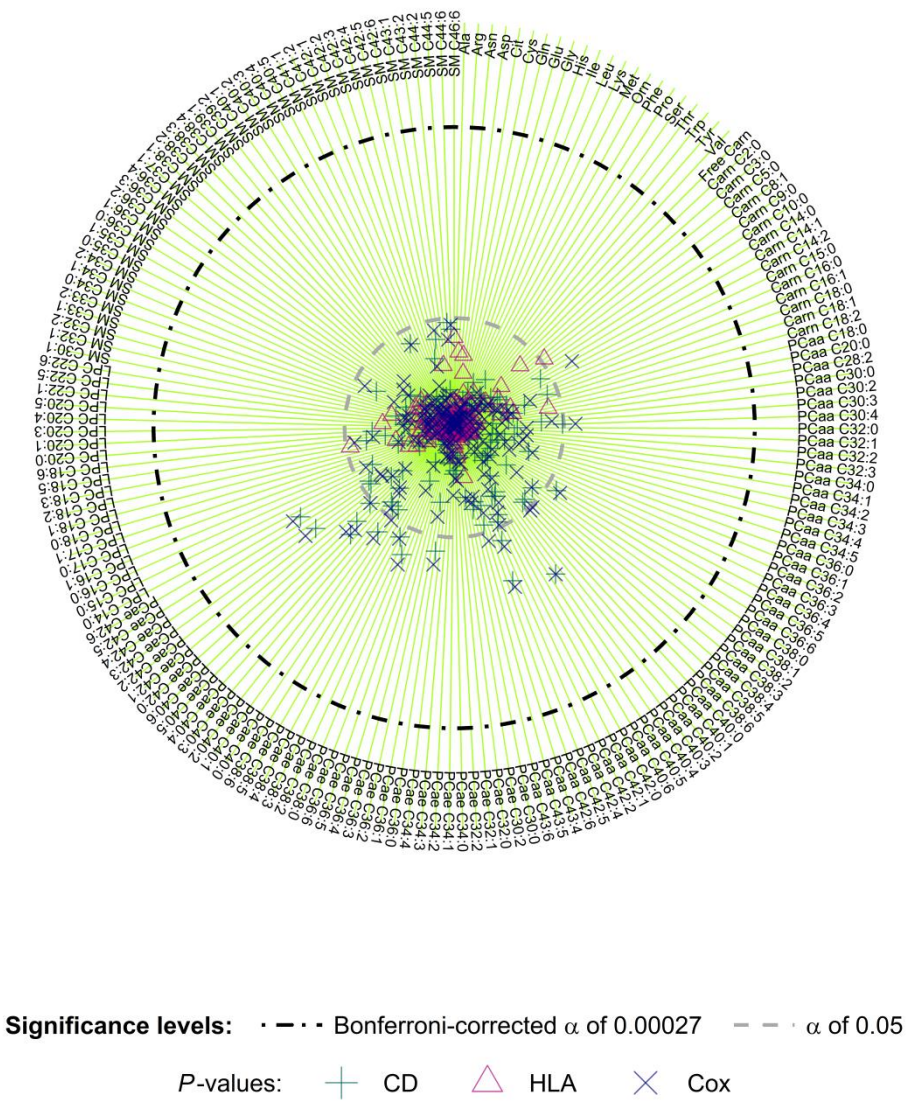

Figure 2: Scatterplot of the sum of diacyl-phosphatidylcholines (PCaa), acyl-alkyl-phosphatidylcholines (PCae), and all phosphatidylcholines (PC) versus the age at diagnosis (if child was diagnosed celiac disease [CD]; cross) or age at analysis (if child was not diagnosed CD up to that date; triangle). Boxplots show the distribution among the groups. Significance for group difference was assessed with linear mixed models adjusted for gender, maternal CD status, and age at blood withdrawal.