Serum proteomics of mother-infant dyads carrying HLA-conferred type 1 diabetes risk

Bhosale, Santosh D.; Moulder, Robert; Suomi, Tomi; Ruohtula, Terhi; Honkanen, Jarno; Virtanen, Suvi M.; Ilonen, Jorma; Elo, Laura L.; Knip, Mikael; Lahesmaa, Riitta

Published in:
iScience

DOI:
10.1016/j.isci.2024.110048

Published: 21/06/2024

Document Version
Final published version

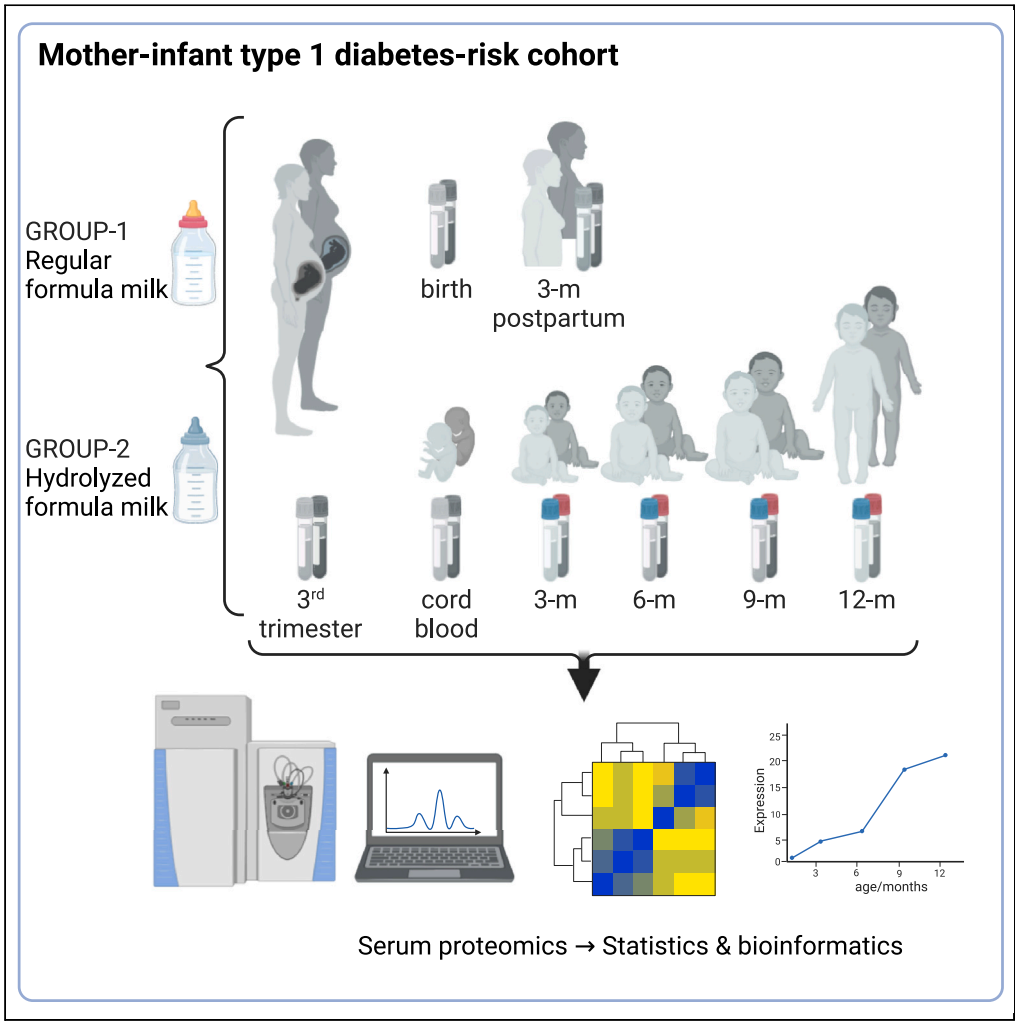
Document License
CC BY-NC

iScience

CellPress
OPEN ACCESS

Article

Serum proteomics of mother-infant dyads carrying HLA-conferred type 1 diabetes risk



Santosh D. Bhosale, Robert Moulder, Tomi Suomi, ..., Laura L. Elo, Mikael Knip, Riitta Lahesmaa

rilahes@utu.fi

Highlights

Longitudinal serum proteomics data for mother-child dyads were generated

Serum proteomes of regular and hydrolyzed milk formula groups were compared

Relationships between mother-infant serum proteomes were detected

Infant serum protein levels were correlated with intestinal permeability

Bhosale et al., iScience 27, 110048
June 21, 2024
https://doi.org/10.1016/j.isci.2024.110048

iScience

CellPress
OPEN ACCESS

Article

Serum proteomics of mother-infant dyads carrying HLA-conferred type 1 diabetes risk

Santosh D. Bhosale,[1,9,10] Robert Moulder,[1,2,10] Tomi Suomi,[1,2,10] Terhi Ruohtula,[3] Jarno Honkanen,[3] Suvi M. Virtanen,[4,5,6] Jorma Ilonen,[7] Laura L. Elo,[1,2,8] Mikael Knip,[1,3,6] and Riitta Lahesmaa[1,2,8,11,*]

SUMMARY

***In-utero* and dietary factors make important contributions toward health and development in early childhood. In this respect, serum proteomics of maturing infants can provide insights into studies of childhood diseases, which together with perinatal proteomes could reveal further biological perspectives. Accordingly, to determine differences between feeding groups and changes in infancy, serum proteomics analyses of mother-infant dyads with HLA-conferred susceptibility to type 1 diabetes (*n* = 22), weaned to either an extensively hydrolyzed or regular cow's milk formula, were made. The LC-MS/MS analyses included samples from the beginning of third trimester, the time of delivery, 3 months postpartum, cord blood, and samples from the infants at 3, 6, 9, and 12 months. Correlations between ranked protein intensities were detected within the dyads, together with perinatal and age-related changes. Comparison with intestinal permeability data revealed a number of significant correlations, which could merit further consideration in this context.**

INTRODUCTION

Type 1 diabetes (T1D) is the result of an autoimmune directed destruction of the insulin producing β cells in the islets of Langerhans of the pancreas and is the most common endocrine disorder in young children.[1] The major genetic determinants of type 1 diabetes are alleles at the HLA-DRB1 and DQB1 loci, whereby there can be both protective and susceptible haplotypes.[2] With the increasing incidence of T1D,[3,4] research efforts have been directed toward its early prediction and prevention.[5–11] Moreover, since priming of the immune system occurs both *in utero* and in infancy, characterization of the biochemical changes during these early stages of development provides an important reference for the recognition of underlying pathologies. Accordingly, prospective sample collections and accompanying analyses using various omics methods have been used to reveal elements of the developing immune system and signs of disease in children carrying genetic risk for T1D.[11–16] Serum proteomics analysis from among the latter cohorts has demonstrated differences prior to islet autoimmunity and subsequently changing patterns preceding the onset of type 1 diabetes.[17] In addition to aiming to establish predictive markers, these studies have also incorporated interventions, seeking to prevent disease onset and progression. The latter have included nasal and oral insulin, hydrolyzed formula milk,[18] and anti-CD3 monoclonal antibodies.[19]

In the present study, infants with HLA-conferred susceptibility to T1D were weaned to an extensively hydrolyzed formula or to a regular cow's milk based formula. To achieve suitable recruitment to the study, 1,468 expecting biological parent pairs were genotyped to select appropriate HLA-conferred risk groups, as previously described.[20] From these, approximately one-third of the couples met the inclusion criteria, and one-third of the children from that group were eligible for the intervention trial. The effect of this dietary intervention on the development of β cell autoimmunity was determined by regular testing for islet cell autoantibodies. Evaluation of clinical data prior to this publication demonstrated that weaning to the hydrolyzed formula did not reduce the incidence of β cell autoimmunity and risk of T1D.[21] However, subsequent analyses have shown that the hydrolyzed formula decreased the intestinal permeability.[20]

In addition to the analysis of these samples in relation to β cell autoimmunity and T1D, we have conducted proteomics measurements on serum samples collected from the participating mother-child dyads (*n* = 22), thus providing an overview of the maturing serum

[1]Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland
[2]InFLAMES Research Flagship Center, University of Turku, Turku, Finland
[3]Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland
[4]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland
[5]Unit of Health Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland
[6]Center for Child Health Research and Research, Development and Innovation Center, Tampere University Hospital, Tampere, Finland
[7]Immunogenetics Laboratory, Institute of Biomedicine, University of Turku, Turku, Finland
[8]Institute of Biomedicine, University of Turku, Turku, Finland
[9]Present address: Precision Biomarker Laboratories, Cedars-Sinai Medical Center, Beverly Hills CA 90211, USA
[10]These authors contributed equally
[11]Lead contact
*Correspondence: rilahes@utu.fi
https://doi.org/10.1016/j.isci.2024.110048

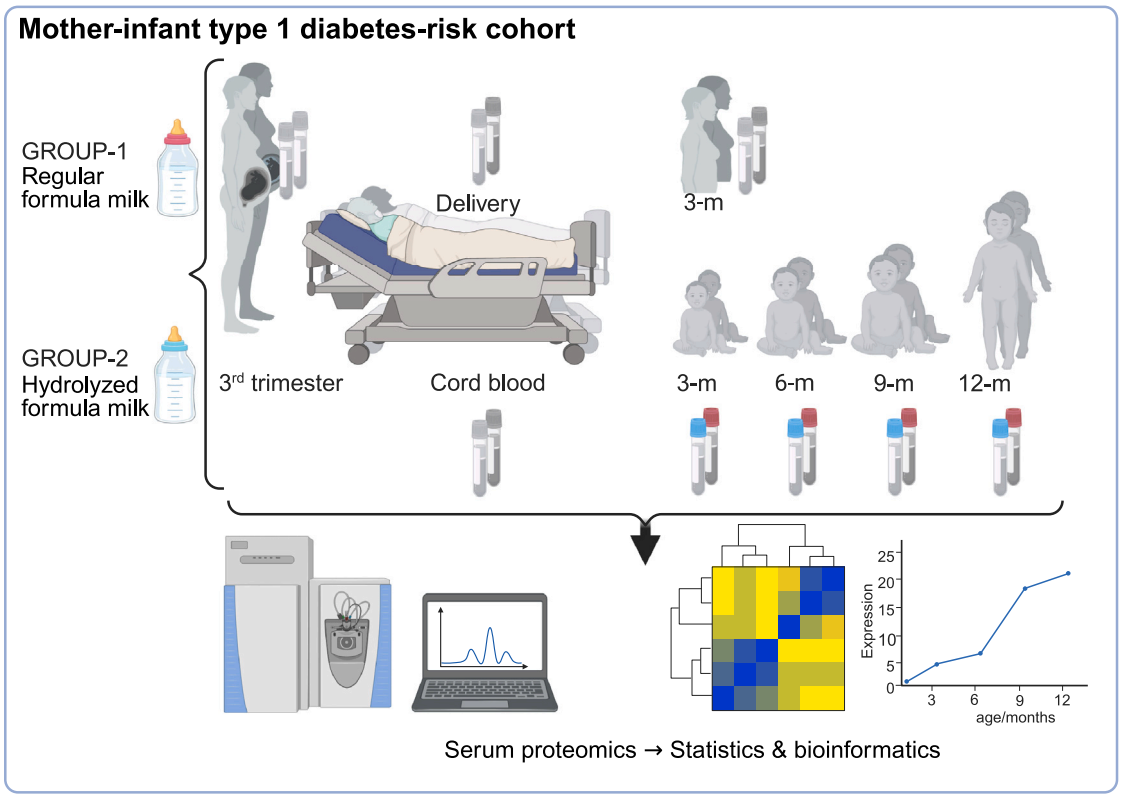


Figure 1. A schematic representation of the study

Serum proteomics of samples from a mother-infant cohort, participating in a study on the influence of weaning to an extensively hydrolyzed cow's milk formula on beta-cell autoimmunity and type 1 diabetes (T1D). The infants carried HLA-conferred susceptibility to T1D. Samples were collected from the mothers at the beginning of the third trimester, at delivery, and 3 months postpartum, and from the infants at birth and at 3, 6, 9, and 12 months. Comparisons were made between the feeding groups, sample types, and other clinical characteristics. The figure was created with BioRender.com.

proteome in the two feeding groups and changes in the maternal proteomes (Figure 1). These analyses thereby provide a unique record of the relationship between mother-infant serum proteomes in the important developmental stages from the third trimester through the first year of life of infants, and an evaluation of dietary-related differences, revealing proteins putatively correlated with intestinal permeability.

RESULTS

Contrasts were detected between the maternal serum proteomes and offspring in the first year of life

Comparison of protein levels in the maternal sera highlighted changes from the third trimester, to delivery and postpartum, with many proteins noted by their contrasting presence/absence in the samples from the different sampling times. Among those that were scarcely detected in the postpartum data were oxytocinase (leucyl-cystinyl aminopeptidase, LNPEP) and the hormones proteins, chorionic somatomammotropin (CSH2) and chorionic gonadotrophins (the CGB7, CGB2, and CGB1 subunits were not distinguished). Similar patterns were observed for pregnancy-specific glycoproteins (PSGs) (Figure 2).

The separate analysis of the cord blood revealed contrasts in the proteins detected. For example, the PSGs, CSH2, and CGB1 were barely detected in the cord blood, with a similar trend observed for the acute phase proteins SAA2 and CRP (Figure 2). In contrast to these, there were a number of proteins that were specific to cord blood. For instance, hemoglobin subunit zeta (HBZ) and hemoglobin subunit gamma-1 (HBG1) were detectable in cord blood but not in the sera from the mothers. Prolactin, alpha-enolase, several collagen proteins (COL1A1 and COL6A1), peptidyl-prolyl *cis*-trans isomerase A, contactin-1, cystatin-M, and alpha-fetoprotein (AFP) were likewise distinguished by their cord blood detection.

The concurrent measurement of the proteomes of the dyads further highlighted proteins detected in the maturing children and infrequently detected in the sera from the mothers (Figure 2). As well as several of the former examples that were marked by their detection in cord blood (i.e., HBG1 and COL1A1), others, similarly noted, included periostin (POSTN), collagen alpha-1(II) chain (chondrocalcin, COL2A1), L-lactate dehydrogenase B chain (LDHB), and neural cell adhesion molecule 1 (NCAM1).

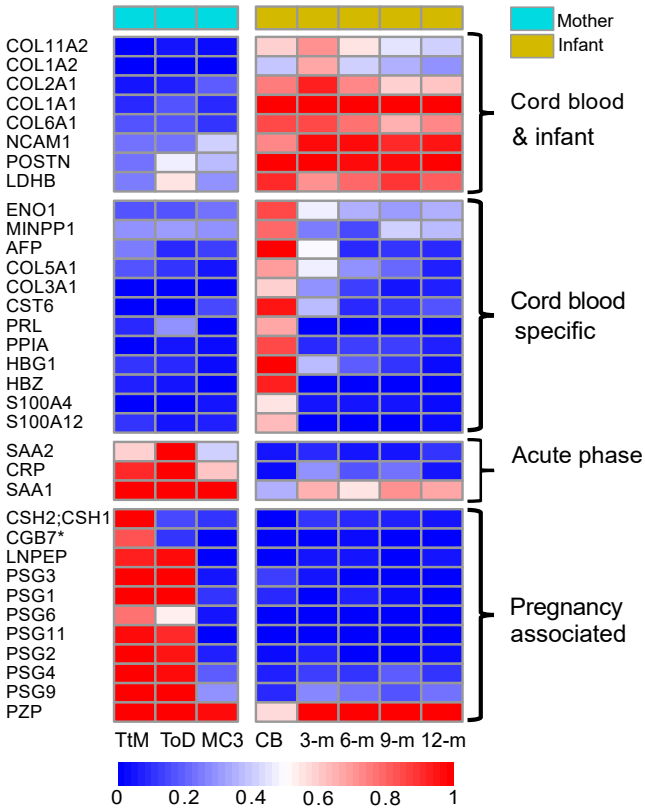


Figure 2. Comparison of the samples from the offspring and mothers emphasized contrasts in the detection of proteins specific to pregnancy, cord blood, and the maturing infants

Examples are shown for the proteins with prominent differences in detection between mothers, infants, and cord blood (n = 22). The scale is shown in relation to the proportion of the subjects, i.e., mothers and offspring, in whom the proteins were detected (1 = all samples, 0 = none). The heatmap shows selected examples that are subdivided into groups on the basis the observed specificity (e.g., cord blood specific or pregnancy associated). The samples types are indicated with the following abbreviations: TtM, ToD, and MC3, for the mothers at the third trimester, time of delivery, and 3 months post-partum, respectively. CB, 3 months, 6 months, 9 months, and 12 months for cord blood and the infants at 3 months, 6 months, 9 months, and 12 months, respectively. For the chorionic gonadotrophins, the CGB7, CGB2, and CGB1 subunits were not distinguished, and are denoted by *CGB7 for brevity. See also Figure S3.

Similarities were revealed between the serum protein levels of mother and child

The proteomics analysis of samples from these dyads facilitated the identification of protein expression patterns transferred from the mother and subsequently sustained in the child. Paired correlation analyses of the ranked expression of the proteins revealed significant correlations (false discovery rate (FDR) <0.05) for 10 proteins (Table 1). These included proteins involved in the complement system (CFHR1 and FCN3) and coagulant/clotting proteins (SERPINA1, FGA, and HABP2), as well as APOA2, lipopolysaccharide-binding protein (LBP), and three immunoglobulin kappa chain proteins. The correlation plots for these are exemplified in the Electronic supplementary material (ESM) Figure S1. The correlations of these proteins between the different sample groups are further represented as a heatmap in Figure 3, from which the consistent mother-to-infant correlation of CFHR1 and the immunoglobulin kappa chain proteins IGKV1-13 and IGKV1D-13 are prominent. On the basis of the detected peptides, it was not possible to establish whether the association was with one or both of the latter two immunoglobulins.

Correlations were detected between the serum proteomes and intestinal inflammation and permeability during the first year of life

Lactulose-mannitol (L/M) tests were performed to assess the intestinal permeability.[22] Comparison of these with the proteomics data revealed that five proteins were significantly correlated with the L/M ratio during the first year (combining data from 3, 6, 9, and 12 months, FDR <0.05). These included fibrinogen and the immunoglobulin variable chain protein IGHV3-72, which were both positively correlated, and APOF, TFRC, and GPX3 that were inversely correlated (Table 2). There were, however, no clear differences between the L/M correlations of the two formula groups.

Human fecal beta-defensin-2 (HBD-2) and calprotectin (FC) were measured from stool samples as markers of gastrointestinal inflammation.[23,24] Correlation analysis was carried out to investigate associations between these and the proteomics data. Although there were no

Table 1. Paired Spearman Rank correlation analyses of the ranked expression of the proteins between the dyads revealed significant correlations for 10 proteins

Protein name	Gene	UniProt i.d.	FDR	r_s
Complement factor H-related protein 1	CFHR1	Q03591	0.0160	0.788
Lipopolysaccharide-binding protein	LBP	P18428	0.0295	0.740
Immunoglobulin kappa variable 1–13, Immunoglobulin kappa variable 1D-13	IGKV1-13, IGKV1D-13	P0DP09, A0A0B4J2D9	0.0295	0.772
Immunoglobulin kappa light chain	*NA*	P0DOX7	0.0295	0.722
Ficolin-3	FCN3	O75636	0.0295	0.719
Alpha-1-antitrypsin	SERPINA1	P01009	0.0295	0.713
Immunoglobulin kappa variable 2–28, Immunoglobulin kappa variable 2D-28	IGKV2-28, IGKV2D-28	A0A075B6P5, P01615	0.0295	0.882
Apolipoprotein A-II	APOA2	P02652	0.0315	0.701
Fibrinogen alpha chain	FGA	P02671	0.0384	0.687
Hyaluronan-binding protein 2	HABP2	Q14520	0.0466	0.674

The proteins are listed with the associated correlation coefficient (r_s) and false discovery rate (all FDRs <0.05). See also Figure S1.

associations consistently significant across all the time points, there were some weakly significant associations with FC (FDR<0.10), as indicated in ESM Table S1. Two of these were at 6 months, IGLV1-40 (positive correlation) and actins (negative), potentially associated with the introduction of solid foods.

Differences and associations of the serum proteome resulting from different study formulas

Analysis of the differences in protein levels from the series of samples from 22 children, revealed a panel of putative markers that were associated with the formula groups (see ESM Table S2). With follow-up analysis of the 9-month samples from an additional 40 children, a panel of nine markers was common to the two sample sets, including PON3, GGH, and several immunoglobulins light chain molecules. However, none of these were significant after FDR correction, and re-analysis of all the 9-month samples using targeted mass spectrometry did not support a significant difference at 9 months (n = 26 vs. 36, hydrolyzed and regular formula, respectively).

A linear mixed-effects model (LME) analysis (n = 22) indicated an interaction effect between age and the groups ($p < 0.05$), such that after 12 months the levels of GGH, CD14, and APOM were higher in the extensively hydrolyzed formula group (Figure 4). In contrast, the immunoglobulin variable kappa chain, IGKV2-30, was less abundant in the extensively hydrolyzed formula group. The interactions were not, however, significant after FDR correction.

Age is a driving factor for the serum levels of many proteins

In addition to the interaction effects, LME analysis defined significant age-related trends for more than fifty proteins (FDR <0.05). Examples of proteins with distinct temporal trends in the first year of life included the increase of several immunoglobulins and a decrease in PI16 (as exemplified in ESM Figure S2). The complete summary of the LME results is shown in ESM Table S3.

DISCUSSION

Pregnancy associated changes in the serum proteome from the third trimester to delivery and postpartum

Proteomics analysis of maternal serum has previously been applied to identify deviations from normal pregnancy and birth defects.[25,26] To tackle the complexity of the serum proteome, both targeted and untargeted mass spectrometry-based approaches have been used, and more recently aptameric-based and proximity extension assays.[27–32] In the current study, using data-dependent mass spectrometry, our data revealed characteristic profiles of healthy deliveries, as detailed with samples from the third trimester, at birth, and then 3 months post-partum.

The discerned differences in detection included several well-established pregnancy-associated proteins (Figure 2). Among these was oxytocinase (LNPEP), which increases in pregnancy to maintain homeostasis and degrades the peptide hormone oxytocin.[33] Similarly detected, chorionic somatomammotropin hormones 1 and 2 (CSH1 and CSH2) are only produced during pregnancy and stimulate lactation, fetal growth, and metabolism.[34] Pregnancy-specific glycoproteins (PSGs) are synthesized in the placenta and become detectable in sera during the later stages of pregnancy.[35] Accordingly, the expression pattern of this group of PSGs was highly specific to the samples collected during pregnancy and at delivery.[33] Another contrast in the detection of pregnancy-associated proteins included pregnancy zone protein, which is a metalloprotease inhibitor that participates in maintaining the immune tolerance of the fetus.[36] Although detected throughout the samples, the relative intensity ranged in the order of several hundred thousand (see ESM Figure S3).

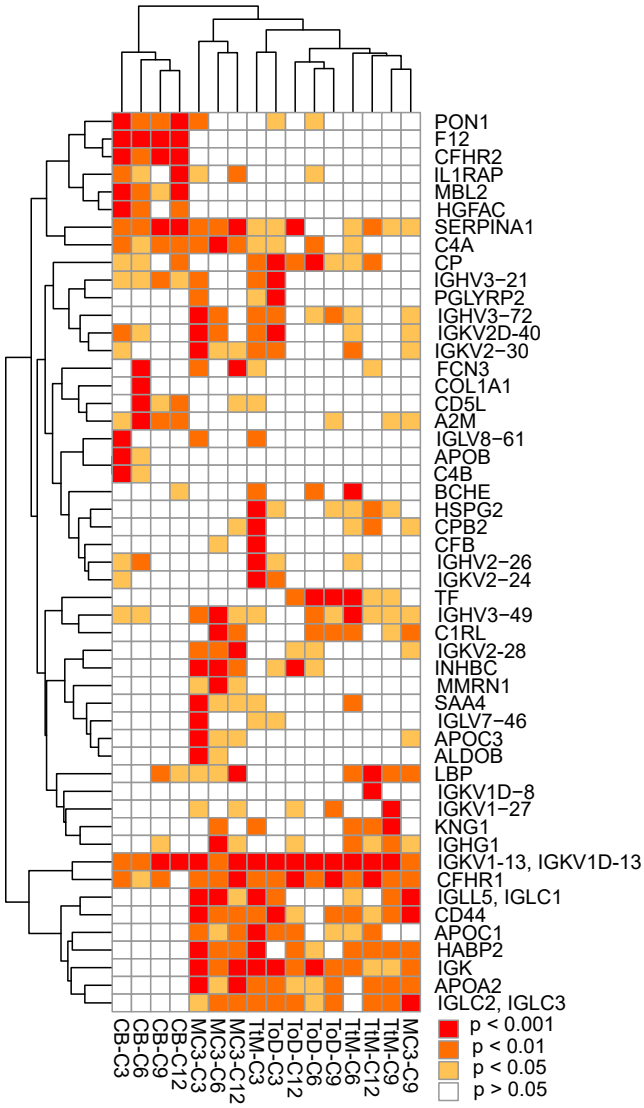


Figure 3. Proteins correlated between the dyads

Comparison of the ranked protein intensities for the different sample draws revealed statistically significant correlations between the temporal samples from the mother-infant dyads (FDR <0.05). The data are shown as a heatmap of the Spearman Rank correlation coefficients. The comparison includes samples from 22 mother-infant dyads.

Proteins specific and excluded from cord blood sera proteome

Although the cord blood sera represent the most direct link between the mother and child, the placenta provides a physical separation between the two. Differences in detection, i.e., presence or absence, reflecting the controlling role of the latter were thus observed. For example, the pregnancy-specific glycoproteins detected in the maternal serum at the time of delivery were not detected in cord blood. Similarly, although a peak in the acute phase proteins SAA2 and CRP was observed in the sera from the mothers at the time of delivery, these were not observed in cord blood. Previous reports have shown that acute phase proteins are not transferred in the context of healthy deliveries.[37,38] In the counter situation, i.e., proteins specific to the child, comparable differences in detection were observed. For example, multiple inositol polyphosphate phosphatase (MINPP1) is associated with bone development and was predominantly detected in cord blood. During fetal development and maturation, there are changes in the structure and composition of the hemoglobin molecules, and HBZ was only detectable in cord blood.[39] In a similar manner, HBG1 is predominant at birth and mostly replaced in the first year of childhood. The detection of prolactin (PRL) was generally specific to the cord blood sera. Prolactin promotes lactation of the mammary gland and has been used to evaluate the threat of pregnancy complications.[40]

A number of proteins potentially transferred from the placenta were detected in cord blood sera (Figure 2). For example, S100A12 has been reported to be present in amniotic fluid and plays a significant anti-infectious and antibacterial role,[41] while S100A4 has been implicated

Table 2. Spearman Rank correlation analysis of the proteomics data and lactulose/mannitol ratios at the different ages revealed several significant correlations

Accession	r_s-3m	r_s-6m	r_s-9m	r_s-12m	Meta FDR	Gene	Protein name
A0A0B4J1Y9	0.47	0.80	0.45	0.51	<0.001	IGHV3-72	Immunoglobulin heavy variable 3-72
P02671	NS	0.57	0.52	0.61	<0.001	FGA	Fibrinogen alpha chain
P22352	−0.54	−0.51	−0.44	−0.52	<0.001	GPX3	Glutathione peroxidase 3
P02786	−0.46	−0.53	−0.44	−0.56	<0.001	TFRC	Transferrin receptor protein 1
Q13790	−0.50	−0.53	−0.49	NS	<0.001	APOF	Apolipoprotein F

The correlation coefficients (r_s) are tabulated together with the *meta*-FDR from the combined time points (NS = not significant).

in calcium uptake to the placenta during pregnancy. Enolase (ENO1) is expressed in the embryo and involved in ontogenesis.[42] Collagen proteins are an integral component of the fetal-maternal interface and are expressed in the trophoblast and decidua. In particular, detection of collagen type V-alpha 1-chain protein (COL5A1) could be representative of this connection. Alpha-fetoprotein (AFP) was distinguished by its detection in cord blood and early samples. AFP produced in the liver of the developing fetus has been assessed during the second trimester of the pregnancy as an indicator for the risk of genetic deviations and birth defects.

Similarities in the serum protein levels between mother and child

Comparisons within the dyads, on the basis of correlation analyses of the ranked protein levels, revealed statistically significant associations for several proteins. These were interpreted to represent traits transferred from the mothers. For example, in keeping with the concept of naturally acquired passive immunity and the linkage between the immune systems of the mother and child, these included several immunoglobulin chain proteins, thus following as a result of immunoglobulin transfer during pregnancy and breastfeeding.[43] In particular, the correlation of the immunoglobulin kappa variable chain proteins (IGKV1-13 and IGKV1D-13), was consistent throughout the temporal samples.

Lipopolysaccharide-binding protein (LBP) is involved in innate immunity and response to gram-negative bacteria. The observed mother to child correlation with LBP levels could similarly reflect immunological cross-over between the mother and child. This is in keeping with our earlier observation of *in utero* priming of the immune system.[44]

Among the other maternally correlated proteins, SERPINA1, FGA, and HABP2 (hyaluronan-binding protein 2) are associated through their roles in coagulation and clotting. HABP2 cleaves fibrinogen to prevent clotting, and SERPINA1 is serine protease that has affinity for thrombin, which initiates FGA cleavage, leading to clotting. As clotting disorders are inheritable, the pattern of expression of this sub-group could be congenital. Similarly, while CFHR1 is involved in complement activation and ficolin-3 (FCN3) is associated with activation of the lectin complement pathway, the association could reflect familiar characteristics in the complement system.[45]

Proteomics changes in the first year of life

Previous studies have reported examples of age-related changes in the plasma and serum proteome in children.[14,16,46,47] In our earlier study of children from birth until the age of 3 years,[16] many changes were observed in the plasma samples during the first year. In keeping with that observation, there were clear age-associated differences between the samples, including many immunoglobulins. Incidentally, in contrast with our earlier study, no depletion step was included in the current analyses, whereas the former incorporated immuno-depletion of the 12 most abundant proteins. The latter targets included IgG, IgA, IgM, IgD, IgE, kappa, and lambda light chains, which in the current study revealed significant positive correlations with age. These trends are a likely result of exposure to dietary antigens and development of the immune system. In terms of decreasing expression, the clearest examples were demonstrated by PI16 and APOA4 (see ESM Figure S2).

Viewing these changes in the context of studies aiming to find early markers of type 1 diabetes, the separation of age-related trends from disease-related changes requires appropriate study design and modeling. This can be especially challenging with the early presentation of islet autoimmunity (<2 years), where the coincident changes due to maturation in the first years are numerous. In this setting, previous studies have been assisted by comparison with age, sex, and risk matched individuals.[11,47] Recent publications, analyzing the changes in serum protein levels from cohorts with HLA-confirmed risk, have demonstrated decreased complement proteins prior to islet autoimmunity and disruptions of the complement system accompanying the pathway to type 1 diabetes onset.[17] Notably, in view of this important observation, age-related changes have been demonstrated for many complement proteins,[47] emphasizing the importance of age considerations in experimental design.[16,47] Among the proteins noted by their contrasting expression between the offspring and the mothers, many were associated with the extracellular matrix and extracellular matrix interaction. These included periostin (POSTN), neural cell adhesion molecule 1 (NCAM1), and several collagens. Periostin is an extracellular matrix protein that is involved in tissue development and regeneration,[48] and NCAM1 is a cell adhesion molecule that is operative in cell-cell and cell to matrix interactions in the developing nervous system. Earlier studies of reference values for NCAM1 in serum have indicated its greater abundance in young children, relative to adults.[49]

In terms of growth, the height at 12 months was positively correlated with COL6A3. COL6A3 is involved in cell adhesion and is important in organizing extracellular matrix proteins. The correlation and latter role could account for this association.

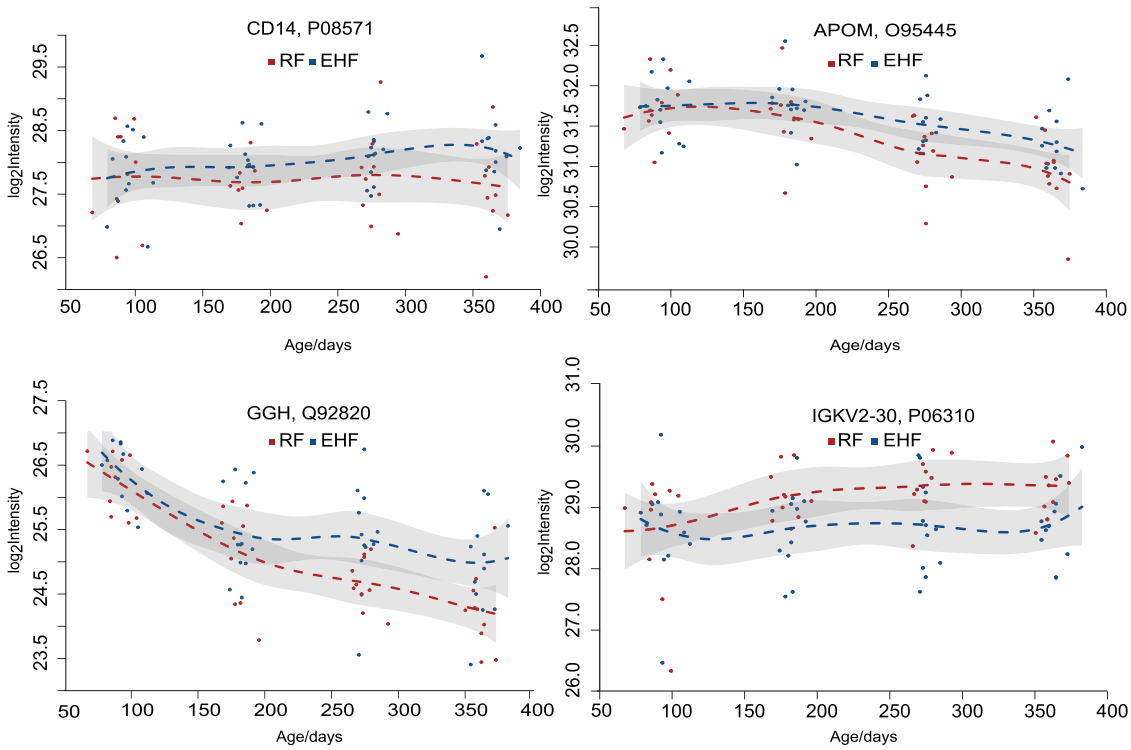


Figure 4. Linear mixed effects analysis revealed interactions between the formula types

The examples shown demonstrate that while there were no differences in the serum protein levels at 3 months, divergence between the groups was observed at 12 months. Although significant *p* values were observed ($p < 0.05$), these were not significant after FDR correction. Red = regular (RF), blue = extensively hydrolyzed (EHF).

Interactions and differences between the formulas and the serum proteome

By the age of 9 months, all of the infants had been exposed to one of the study formulas, and an analysis was conducted at this time point comparing the two groups ($n = 26$ vs. 36, hydrolyzed and regular formula, respectively). From this comparison, lower levels of serum paraoxonase/lactonase 3 (PON3) and gamma-glutamyl hydrolase (GGH) were detected in the regular formula group compared to the hydrolyzed formula group. However, the targeted mass spectrometry verification measurements in the same samples did not support these differences, potentially due to the loss of statistical power relative to the longitudinal comparison. Nevertheless, from the LME analysis of the data from the time series used for discovery ($n = 10$ vs. 12), several interactions were apparent as displayed by diverging group-wise differences in the levels of GGH, APOM, CD14, and IGKV2-30 at 12 months (Figure 4).

In the context of milk and formula composition, GGH hydrolyses pteroylpolyglutamate, a major folate component of milk.[50,51] Potentially, the detected interaction could arise from the dietary difference and the influence of the latter process upon the serum levels of GGH. Regarding the interaction for CD14, dysfunction of the gut barrier has been previously demonstrated to be associated with higher plasma concentrations of CD14.[52] Further studies would be needed to establish other implications of this observation. APOM is mostly associated with high-density lipoproteins and involved in lipid transport.[53] Among the latter, APOM is the main carrier of plasma sphingosine-1-phosphate (S1P)[54] and APOM-bound S1P is involved in the maintenance of vascular integrity,[55] as has been reported in relation to S1P and the inflammatory bowel disease ulcerative colitis. The associations of APOM and S1P suggest that these could be relevant in the current example, whereby the difference in serum APOM (higher in the extensively hydrolyzed formula group) could reflect better integrity of the intestinal epithelium and the lower intestinal permeability in the group that received extensively hydrolyzed formula, as reported from the earlier analysis of L/M data from this study.[20] In addition, characterization of the serum lipidome from this cohort has revealed that sphingomyelins were higher in the extensively hydrolyzed formula group as compared to the regular formula group.[56] However, a straightforward connection between the balance of these with S1P and serum APOM is not clear, especially since decreased sphingomyelin has been reported from measurements with the overexpression of APOM in mouse.[57] In the context of potential markers and associations with type 1 diabetes, increased serum APOM has been detected in type 1 diabetes susceptible children after seroconversion[13] and type 1 diabetes risk has been associated with SNPs of APOM.[58]

Serum protein correlations with intestinal permeability

Although the study formula did not prevent β cell autoimmunity, subsequently published analysis of the lactulose to mannitol ratio (L/M ratio) measurements at 9 months of age revealed decreased gut permeability in the infants fed extensively hydrolyzed formula. This difference was

taken as being a potential outcome of reduced exposure to dietary antigens in the extensively hydrolyzed formula group. At this stage, the beneficial or adverse effects of this difference have yet to be established.[20] Nevertheless, the importance of intestinal permeability in maintaining a barrier against potential triggers of type 1 diabetes is a relevant concern,[59,60] and yet the potential of dietary prevention remains to be successfully demonstrated.

Comparison of the L/M ratios with the proteomics data revealed a range of positive and inverse correlations, none of which, however, were divergent between the formula groups (Table 2). Notably, a high L/M ratio is indicative of increased permeability, suggesting that the positive correlation reflects a more adverse situation. For example, the immunoglobulin IGHV3-72 was positively correlated with the L/M ratio, which could be the result of differences in the antigenic challenge and the diminished integrity of the intestinal barrier. Fibrinogen, which was also positively correlated, is well known for its role in coagulation and wound healing, and is upregulated after bleeding or clotting events.[61] The observed correlation with the intestinal permeability suggests that serum fibrinogen levels could even reflect the balance of intestinal permeability.[62]

APOF, TFRC, and GPX3 were all inversely correlated with the L/M ratios (FDR <0.05). APOF primarily associates with LDL and has been linked with inhibition of cholesteryl ester transfer protein (CETP), thus affecting cholesterol transport.[63] In this respect, although cholesterol homeostasis is mostly controlled by the liver, the intestine also plays a role and can affect plasma cholesterol.[64] Among other processes, transferrin receptor protein 1 (TFRC) is involved in iron transport, and the uptake of TFRC-bound iron is the main pathway for iron entry into cells. Severe anemia has previously been associated with increased intestinal permeability to dietary macromolecules and bacteria.[65] The inverse correlation observed suggests that insufficient or reduced serum TFRC might be related to changes in intestinal permeability. GPX3 catalyzes the reduction of peroxides by glutathione and protects cells and enzymes from oxidative damage.[66] An interpretation of this association, which suggests lower GPX3 is observed when permeability is increased, suggests that the protection against oxidative stress might be compromised in such circumstances. Together, these correlated proteins may reflect how a balance of proteins involved in lipid and iron transport, together with protection against oxidative damage, are important in maintaining intestinal permeability. On the basis of these observations, future studies could take into consideration serum levels of TFRC, APOF, and GPX3 in relation to intestinal permeability. Although further studies would be required to verify these observations, their accessibility could prove useful if confirmed, especially in the light of the growing incidence of autoimmune diseases and concerns of the role of the gut.

Limitations of the study

These analyses of non-depleted serum were limited by the depth of coverage of the serum proteomes. However, co-analysis of reference samples, from which the most abundant proteins had been depleted, was employed to facilitate detection of less abundant proteins and help track their occurrence in the non-depleted samples. The use of depletion or enrichment for all samples, using a conventional or a nanoparticle-based approach, respectively, would have expanded the number of proteins detected. Furthermore, the analysis of a larger set of the samples during discovery and validation in the full sample time series would have been advantageous. In particular, in view of the capabilities and developments with data independent mass spectrometry and use of higher throughput robust separations systems, future studies could provide better coverage of more samples in a time efficient manner.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Study formula
- METHOD DETAILS
  - Sample preparation
  - Data-dependent LC-MS/MS analysis
  - Targeted LC-MS/MS analysis
  - Data analysis – LC-MS/MS
  - The lactulose–mannitol test
  - Analysis of fecal calprotectin and HBD-2
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Linear Mixed Effects Model
  - Correlation Analysis
  - The influence of formula type on serum proteins
  - Differences between sample types

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110048.

ACKNOWLEDGMENTS

The work was financially supported by Juvenile Diabetes Research Foundation Ltd, the Academy of Finland, the Finnish Diabetes Foundation, the Sigrid Jusélius Foundation, and InFLAMES Flagship Program.

R.L. received funding from the Academy of Finland (grants 292335, 294337, 319280, 31444, 319280, 329277, and 331790), Business Finland and by grants from the JDRF, Sigrid Jusélius Foundation, Jane and Aatos Erkko Foundation, Finnish Diabetes Foundation, and the Finnish Cancer Foundation. L.E. reports grants from the European Research Council ERC (677943), Academy of Finland (296801, 310561, 314443, 329278, 335434, and 335611), and Sigrid Jusélius Foundation during the conduct of the study. R.L. and M.K. were supported by the Academy of Finland, AoF, Centre of Excellence in Molecular Systems Immunology and Physiology Research (2012–2017, grants 250114 and 292482). M.K. also received support from the Sigrid Jusélius Foundation, Helsinki University Hospital Research Funds and the Liv and Hälsa Fund.

Research at Turku Bioscience Center (L.E. and R.L.) was supported by University of Turku Graduate School (UTUGS), Biocenter Finland, ELIXIR Finland, and the InFLAMES Flagship Program of the Academy of Finland (decision number: 337530).

The work presented in this paper was conducted at Turku Bioscience Center (University of Turku and Åbo Akademi University). The proteomics analysis were conducted at the Turku Bioscience Proteomics core, which is supported by Biocenter Finland.

AUTHOR CONTRIBUTIONS

Conceptualization, M.K. and R.L.; methodology, S.D.B., R.M., and T.S.; investigation, S.D.B., R.M., T.S., T.R., J.H., S.V., and J.I.; writing – original draft, R.M. and S.D.B.; writing – review and editing, all authors reviewed and revised the manuscript and approved the submitted version; funding acquisition, M.K., R.L., and L.E.; resources, M.K. and S.V.; supervision, R.L., M.K., and L.E.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 24, 2023
Revised: March 22, 2024
Accepted: May 17, 2024
Published: May 22, 2024

REFERENCES

1. Atkinson, M.A. (2012). The pathogenesis and natural history of type 1 diabetes. Cold Spring Harb. Perspect. Med. *2*, a007641. https://doi.org/10.1101/cshperspect.a007641.
2. Hermann, R., Turpeinen, H., Laine, A.P., Veijola, R., Knip, M., Simell, O., Sipilä, I., Akerblom, H.K., and Ilonen, J. (2003). HLA DR-DQ-encoded genetic determinants of childhood-onset type 1 diabetes in Finland: an analysis of 622 nuclear families. Tissue Antigens *62*, 162–169. https://doi.org/10.1034/j.1399-0039.2003.00071.x.
3. Schmidt, C.W. (2011). Questions persist: environmental factors in autoimmune disease. Environ. Health Perspect. *119*, A249–A253. https://doi.org/10.1289/ehp.119-a248.
4. Norris, J.M., Barriga, K., Klingensmith, G., Hoffman, M., Eisenbarth, G.S., Erlich, H.A., and Rewers, M. (2003). Timing of initial cereal exposure in infancy and risk of islet autoimmunity. JAMA *290*, 1713–1720. https://doi.org/10.1001/jama.290.13.1713.
5. Kukko, M., Virtanen, S.M., Toivonen, A., Simell, S., Korhonen, S., Ilonen, J., Simel, O., and Knip, M. (2004). Geographical variation in risk HLA-DQB1 genotypes for type 1 diabetes and signs of beta-cell autoimmunity in a high-incidence country. Diabetes Care *27*, 676–681. https://doi.org/10.2337/diacare.27.3.676.
6. Skyler, J.S. (2007). Prediction and Prevention of Type 1 Diabetes: Progress, Problems, and Prospects. Clin. Pharmacol. Ther. *81*, 768–771. https://doi.org/10.1038/sj.clpt.6100179.
7. Oresic, M., Gopalacharyulu, P., Mykkänen, J., Lietzen, N., Mäkinen, M., Nygren, H., Simell, S., Simell, V., Hyöty, H., Veijola, R., et al. (2013). Cord serum lipidome in prediction of islet autoimmunity and type 1 diabetes. Diabetes *62*, 3268–3274. https://doi.org/10.2337/db13-0159.
8. Kallionpaa, H., Elo, L.L., Laajala, E., Mykkanen, J., Ricano-Ponce, I., Vaarma, M., Laajala, T.D., Hyoty, H., Ilonen, J., Veijola, R., et al. (2014). Innate immune activity is detected prior to seroconversion in children with HLA-conferred type 1 diabetes susceptibility. Diabetes *63*, 2402–2414. https://doi.org/10.2337/db13-1775.
9. Moulder, R., Bhosale, S.D., Lahesmaa, R., and Goodlett, D.R. (2017). The progress and potential of proteomic biomarkers for type 1 diabetes in children. Expert Rev. Proteomics *14*, 31–41. https://doi.org/10.1080/14789450.2017.1265449.
10. Herold, K.C., Bundy, B.N., Long, S.A., Bluestone, J.A., DiMeglio, L.A., Dufort, M.J., Gitelman, S.E., Gottlieb, P.A., Krischer, J.P., Linsley, P.S., et al. (2019). An Anti-CD3 Antibody, Teplizumab, in Relatives at Risk for Type 1 Diabetes. N. Engl. J. Med. *381*, 603–613. https://doi.org/10.1056/NEJMoa1902226.
11. Nakayasu, E.S., Bramer, L.M., Ansong, C., Schepmoes, A.A., Fillmore, T.L., Gritsenko, M.A., Clauss, T.R., Gao, Y., Piehowski, P.D., Stanfill, B.A., et al. (2023). Plasma protein biomarkers predict the development of persistent autoantibodies and type 1 diabetes 6 months prior to the onset of autoimmunity. Cell Rep. Med. *4*, 101093. https://doi.org/10.1016/j.xcrm.2023.101093.
12. Moulder, R., Bhosale, S.D., Erkkilä, T., Laajala, E., Salmi, J., Nguyen, E.V., Kallionpää, H., Mykkänen, J., Vähä-Mäkilä, M., Hyöty, H., et al. (2015). Serum proteomes distinguish children developing type 1 diabetes in a cohort with HLA-conferred susceptibility. Diabetes *64*, 2265–2278. https://doi.org/10.2337/db14-0983.
13. von Toerne, C., Laimighofer, M., Achenbach, P., Beyerlein, A., de Las Heras Gala, T., Krumsiek, J., Theis, F.J., Ziegler, A.G., and Hauck, S.M. (2017). Peptide serum markers in islet autoantibody-positive children. Diabetologia *60*, 287–295. https://doi.org/10.1007/s00125-016-4150-x.
14. Liu, C.W., Bramer, L., Webb-Robertson, B.J., Waugh, K., Rewers, M.J., and Zhang, Q. (2016). Temporal profiles of plasma proteome during childhood development. J. Proteomics *152*, 321–328.
15. Liu, C.W., Bramer, L., Webb-Robertson, B.J., Waugh, K., Rewers, M.J., and Zhang, Q. (2018). Temporal expression profiling of plasma proteins reveals oxidative stress in

early stages of Type 1 Diabetes progression. J. Proteomics *172*, 100–110.
16. Lietzen, N., Cheng, L., Moulder, R., Siljander, H., Laajala, E., Harkonen, T., Peet, A., Vehtari, A., Tillmann, V., Knip, M., et al. (2018). Characterization and non-parametric modeling of the developing serum proteome during infancy and early childhood. Sci. Rep. *8*, 5883. https://doi.org/10.1038/s41598-018-24019-5.
17. Webb-Robertson, B.-J.M., Nakayasu, E.S., Dong, F., Waugh, K.C., Flores, J.E., Bramer, L.M., Schepmoes, A.A., Gao, Y., Fillmore, T.L., Onengut-Gumuscu, S., et al. (2024). Decrease in multiple complement proteins associated with development of islet autoimmunity and type 1 diabetes. iScience *27*, 108769. https://doi.org/10.1016/j.isci.2023.108769.
18. Wherrett, D.K. (2014). Trials in the prevention of type 1 diabetes: current and future. Can. J. Diabetes *38*, 279–284. https://doi.org/10.1016/j.jcjd.2014.05.004.
19. Herold, K.C., Bundy, B.N., Long, S.A., Bluestone, J.A., DiMeglio, L.A., Dufort, M.J., Gitelman, S.E., Gottlieb, P.A., Krischer, J.P., Linsley, P.S., et al. (2019). An Anti-CD3 Antibody, Teplizumab, in Relatives at Risk for Type 1 Diabetes. N. Engl. J. Med. *381*, 603–613. https://doi.org/10.1056/NEJMoa1902226.
20. Siljander, H., Jason, E., Ruohtula, T., Selvenius, J., Koivusaari, K., Salonen, M., Ahonen, S., Honkanen, J., Ilonen, J., Vaarala, O., et al. (2021). Effect of Early Feeding on Intestinal Permeability and Inflammation Markers in Infants with Genetic Susceptibility to Type 1 Diabetes: A Randomized Clinical Trial. J. Pediatr. *238*, 305–311.e3. https://doi.org/10.1016/j.jpeds.2021.07.042.
21. for the TRIGR Study Group, W.G., Knip, M., Åkerblom, H.K., Taji, E.A., Becker, D., Bruining, J., Castano, L., Danne, T., de Beaufort, C., Dosch, H.M., et al. (2018). Effect of Hydrolyzed Infant Formula vs Conventional Formula on Risk of Type 1 Diabetes: The TRIGR Randomized Clinical Trial. JAMA *319*, 38–48. https://doi.org/10.1001/jama.2017.19826.
22. Denno, D.M., VanBuskirk, K., Nelson, Z.C., Musser, C.A., Hay Burgess, D.C., and Tarr, P.I. (2014). Use of the Lactulose to Mannitol Ratio to Evaluate Childhood Environmental Enteric Dysfunction: A Systematic Review. Clin. Infect. Dis. *59*, S213–S219. https://doi.org/10.1093/cid/ciu541.
23. Tøn, H., Brandsnes, Dale, S., Holtlund, J., Skuibina, E., Schjønsby, H., and Johne, B. (2000). Improved assay for fecal calprotectin. Clin. Chim. Acta *292*, 41–54. https://doi.org/10.1016/s0009-8981(99)00206-5.
24. Wehkamp, J., Fellermann, K., Herrlinger, K.R., Baxmann, S., Schmidt, K., Schwind, B., Duchrow, M., Wohlschläger, C., Feller, A.C., and Stange, E.F. (2002). Human beta-defensin 2 but not beta-defensin 1 is expressed preferentially in colonic mucosa of inflammatory bowel disease. Eur. J. Gastroenterol. Hepatol. *14*, 745–752. https://doi.org/10.1097/00042737-200207000-00006.
25. Navajas, R., Corrales, F., and Paradela, A. (2021). Quantitative proteomics-based analyses performed on pre-eclampsia samples in the 2004–2020 period: a systematic review. Clin. Proteomics *18*, 6. https://doi.org/10.1186/s12014-021-09313-1.
26. Chen, L., Gu, H., Li, J., Yang, Z.Y., Sun, X., Zhang, L., Shan, L., Wu, L., Wei, X., Zhao, Y., et al. (2016). Comprehensive maternal serum proteomics identifies the cytoskeletal proteins as non-invasive biomarkers in prenatal diagnosis of congenital heart defects. Sci. Rep. *6*, 19248. https://doi.org/10.1038/srep19248.
27. Aghaeepour, N., Lehallier, B., Baca, Q., Ganio, E.A., Wong, R.J., Ghaemi, M.S., Culos, A., El-Sayed, Y.Y., Blumenfeld, Y.J., Druzin, M.L., et al. (2018). A proteomic clock of human pregnancy. Am. J. Obstet. Gynecol. *218*, 347.e1-14. https://doi.org/10.1016/j.ajog.2017.12.208.
28. Hao, S., You, J., Chen, L., Zhao, H., Huang, Y., Zheng, L., Tian, L., Maric, I., Liu, X., Li, T., et al. (2020). Changes in pregnancy-related serum biomarkers early in gestation are associated with later development of preeclampsia. PLoS One *15*, e0230000. https://doi.org/10.1371/journal.pone.0230000.
29. Erez, O., Romero, R., Maymon, E., Chaemsaithong, P., Done, B., Pacora, P., Panaitescu, B., Chaiworapongsa, T., Hassan, S.S., and Tarca, A.L. (2017). The prediction of late-onset preeclampsia: Results from a longitudinal proteomics study. PLoS One *12*, e0181468. https://doi.org/10.1371/journal.pone.0181468.
30. Romero, R., Erez, O., Maymon, E., Chaemsaithong, P., Xu, Z., Pacora, P., Chaiworapongsa, T., Done, B., Hassan, S.S., and Tarca, A.L. (2017). The maternal plasma proteome changes as a function of gestational age in normal pregnancy: a longitudinal study. Am. J. Obstet. Gynecol. *217*, 67.e1.
31. Hedman, A.M., Lundholm, C., Andolf, E., Pershagen, G., Fall, T., and Almqvist, C. (2020). Longitudinal plasma inflammatory proteome profiling during pregnancy in the Born into Life study. Sci. Rep. *10*, 17819. https://doi.org/10.1038/s41598-020-74722-5.
32. Ravnsborg, T., Svaneklink, S., Andersen, L.L.T., Larsen, M.R., Jensen, D.M., and Overgaard, M. (2019). First-trimester proteomic profiling identifies novel predictors of gestational diabetes mellitus. PLoS One *14*, e0214457. https://doi.org/10.1371/journal.pone.0214457.
33. Melander, S.E.J. (1961). Oxytocinase Activity of Plasma of Pregnant Women. Nature *191*, 176–177. https://doi.org/10.1038/191176a0.
34. Josimovich, J.B., Atwood, B.L., and Goss, D.A. (1963). Luteotrophic, Immunologic and Electrophoretic Properties of Human Placental Lactogen. Endocrinology *73*, 410–420. https://doi.org/10.1210/endo-73-4-410.
35. Thompson, J., Koumari, R., Wagner, K., Barnert, S., Schleussner, C., Schrewe, H., Zimmermann, W., Müller, G., Schempp, W., Zaninetta, D., et al. (1990). The human pregnancy-specific glycoprotein genes are tightly linked on the long arm of chromosome 19 and are coordinately expressed. Biochem. Biophys. Res. Commun. *167*, 848–859. https://doi.org/10.1016/0006-291X(90)92103-7.
36. von Schoultz, B. (1974). A quantitative study of the pregnancy zone protein in the sera of pregnant and puerperal women. Am. J. Obstet. Gynecol. *119*, 792–797. 0002-9378(74)90092-1 [pii].
37. Cicarelli, L.M., Perroni, A.G., Zugaib, M., de Albuquerque, P.B., and Campa, A. (2005). Maternal and cord blood levels of serum amyloid A, C-reactive protein, tumor necrosis factor-alpha, interleukin-1beta, and interleukin-8 during and after delivery. Mediat. Inflamm. *2005*, 96–100. https://doi.org/10.1155/MI.2005.96.
38. Phillips, R.J., Heesom, K.J., Trinder, J., and Bernal, A.L. (2014). Human maternal plasma proteomic changes with parturition. EuPA Open Proteomics *5*, 10–20.
39. Chui, D.H., Mentzer, W.C., Patterson, M., Iarocci, T.A., Embury, S.H., Perrine, S.P., Mibashan, R.S., and Higgs, D.R. (1989). Human embryonic zeta-globin chains in fetal and newborn blood. Blood *74*, 1409–1414. S0006-4971(20)82595-6 [pii].
40. Alawad, Z.M., and Al-Omary, H.L. (2019). Maternal and cord blood prolactin level and pregnancy complications. Pak. J. Med. Sci. *35*, 1122–1127. https://doi.org/10.12669/pjms.35.4.558.
41. Buhimschi, I.A., Zhao, G., Pettker, C.M., Bahtiyar, M.O., Magloire, L.K., Thung, S., Fairchild, T., and Buhimschi, C.S. (2007). The receptor for advanced glycation end products (RAGE) system in women with intraamniotic infection and inflammation. Am. J. Obstet. Gynecol. *196*, 181.e1
42. Fougerousse, F., Edom-Vovard, F., Merkulova, T., Ott, M.-O., Durand, M., Butler-Browne, G., and Keller, A. (2001). The muscle-specific enolase is an early marker of human myogenesis. J. Muscle Res. Cell Motil. *22*, 535–544. https://doi.org/10.1023/A:1015008208007.
43. Chirico, G., Marzollo, R., Cortinovis, S., Fonte, C., and Gasparoni, A. (2008). Antiinfective Properties of Human Milk. J. Nutr. *138*, 1801S–1806S. https://doi.org/10.1093/jn/138.9.1801S.
44. Kallionpaa, H., Laajala, E., Oling, V., Harkonen, T., Tillmann, V., Dorshakova, N.V., Ilonen, J., Lahdesmaki, H., Knip, M., Lahesmaa, R., et al. (2014). Standard of hygiene and immune adaptation in newborn infants. Clin. Immunol. *155*, 136–147. https://doi.org/10.1016/j.clim.2014.09.009.
45. Fries, L.F., O'Shea, J.J., and Frank, M.M. (1986). Inherited deficiencies of complement and complement-related proteins. Clin. Immunol. Immunopathol. *40*, 37–49. https://doi.org/10.1016/0090-1229(86)90067-x.
46. Bjelosevic, S., Pascovici, D., Ping, H., Karlaftis, V., Zaw, T., Song, X., Molloy, M.P., Monagle, P., and Ignjatovic, V. (2017). Quantitative Age-specific Variability of Plasma Proteins in Healthy Neonates, Children and Adults. Mol. Cell. Proteomics *16*, 924–935. https://doi.org/10.1074/mcp.M116.066720.
47. Hirvonen, M.K., Lietzén, N., Moulder, R., Bhosale, S.D., Koskenniemi, J., Vähä-Mäkilä, M., Nurmio, M., Orešič, M., Ilonen, J., Toppari, J., et al. (2023). Serum APOC1 levels are decreased in young autoantibody positive children who rapidly progress to type 1 diabetes. Sci. Rep. *13*, 15941. https://doi.org/10.1038/s41598-023-43039-4.
48. Merle, B., and Garnero, P. (2012). The multiple facets of periostin in bone metabolism. Osteoporos. Int. *23*, 1199–1212. https://doi.org/10.1007/s00198-011-1892-7.
49. Glüer, S., Wunder, M.A., Schelp, C., Radtke, E., and Gerardy-Schahn, R. (1998). Polysialylated Neural Cell Adhesion Molecule Serum Levels in Normal Children. Pediatr. Res. *44*, 915–919. https://doi.org/10.1203/00006450-199812000-00015.
50. Galivan, J., Ryan, T., Rhee, M., Yao, R., and Chave, K. (1999). Glutamyl hydrolase: properties and pharmacologic impact. Semin. Oncol. *26*, 33–37.
51. O'Connor, D.L., Tamura, T., and Picciano, M.F. (1991). Pteroylpolyglutamates in human milk. Am. J. Clin. Nutr. *53*, 930–934. https://doi.org/10.1093/ajcn/53.4.930.

52. Tabung, F.K., Birmann, B.M., Epstein, M.M., Martínez-Maza, O., Breen, E.C., Wu, K., and Giovannucci, E.L. (2017). Influence of Dietary Patterns on Plasma Soluble CD14, a Surrogate Marker of Gut Barrier Dysfunction. Curr. Dev. Nutr. *1*, e001396. https://doi.org/10.3945/cdn.117.001396.
53. Li, Y., Zhou, J., Qiu, J., Huang, Z., Wang, W., Wu, P., and Feng, A. (2020). Berberine reduces gut-vascular barrier permeability via modulation of ApoM/S1P pathway in a model of polymicrobial sepsis. Life Sci. *261*, 118460. https://doi.org/10.1016/j.lfs.2020.118460.
54. Christoffersen, C., and Nielsen, L.B. (2012). Apolipoprotein M--a new biomarker in sepsis. Crit. Care *16*, 126. https://doi.org/10.1186/cc11320.
55. Galvani, S., Sanson, M., Blaho, V.A., Swendeman, S.L., Obinata, H., Conger, H., Dahlbäck, B., Kono, M., Proia, R.L., Smith, J.D., and Hla, T. (2015). HDL-bound sphingosine 1-phosphate acts as a biased agonist for the endothelial cell receptor S1P1 to limit vascular inflammation. Sci. Signal. *8*, ra79. https://doi.org/10.1126/scisignal.aaa2581.
56. Lamichhane, S., Siljander, H., Salonen, M., Ruohtula, T., Virtanen, S.M., Ilonen, J., Hyötyläinen, T., Knip, M., and Orešič, M. (2022). Impact of Extensively Hydrolyzed Infant Formula on Circulating Lipids During Early Life. Front. Nutr. *9*, 859627. https://doi.org/10.3389/fnut.2022.859627.
57. Liu, M., Seo, J., Allegood, J., Bi, X., Zhu, X., Boudyguina, E., Gebre, A.K., Avni, D., Shah, D., Sorci-Thomas, M.G., et al. (2014). Hepatic Apolipoprotein M (ApoM) Overexpression Stimulates Formation of Larger ApoM/Sphingosine 1-Phosphate-enriched Plasma High Density Lipoprotein. J. Biol. Chem. *289*, 2801–2814. https://doi.org/10.1074/jbc.M113.499913.
58. Christoffersen, C. (2021). Apolipoprotein M-A Marker or an Active Player in Type II Diabetes? Front. Endocrinol. *12*, 665393. https://doi.org/10.3389/fendo.2021.665393.
59. Bosi, E., Molteni, L., Radaelli, M.G., Folini, L., Fermo, I., Bazzigaluppi, E., Piemonti, L., Pastore, M.R., and Paroni, R. (2006). Increased intestinal permeability precedes clinical onset of type 1 diabetes. Diabetologia *49*, 2824–2827. https://doi.org/10.1007/s00125-006-0465-3.
60. Mønsted, M.Ø., Falck, N.D., Pedersen, K., Buschard, K., Holm, L.J., and Haupt-Jorgensen, M. (2021). Intestinal permeability in type 1 diabetes: An updated comprehensive overview. J. Autoimmun. *122*, 102674. https://doi.org/10.1016/j.jaut.2021.102674.
61. Vilar, R., Fish, R.J., Casini, A., and Neerman-Arbez, M. (2020). Fibrin(ogen) in human disease: both friend and foe. Haematologica *105*, 284–296. https://doi.org/10.3324/haematol.2019.236901.
62. Seltana, A., Cloutier, G., Reyes Nicolas, V., Khalfaoui, T., Teller, I.C., Perreault, N., and Beaulieu, J.-F. (2022). Fibrin(ogen) Is Constitutively Expressed by Differentiated Intestinal Epithelial Cells and Mediates Wound Healing. Front. Immunol. *13*, 916187. https://doi.org/10.3389/fimmu.2022.916187.
63. Liu, Y., and Morton, R.E. (2020). Apolipoprotein F: a natural inhibitor of cholesteryl ester transfer protein and a key regulator of lipoprotein metabolism. Curr. Opin. Lipidol. *31*, 194–199. https://doi.org/10.1097/MOL.0000000000000688.
64. Kruit, J.-K., Groen, A.K., van Berkel, T.J., and Kuipers, F. (2006). Emerging roles of the intestine in control of cholesterol metabolism. World J. Gastroenterol. *12*, 6429–6439. https://doi.org/10.3748/wjg.v12.i40.6429.
65. Cabantchik, Z.I. (2014). Labile iron in cells and body fluids: physiology, pathology, and pharmacology. Front. Pharmacol. *5*, 45. https://doi.org/10.3389/fphar.2014.00045.
66. Chang, C., Worley, B.L., Phaëton, R., and Hempel, N. (2020). Extracellular Glutathione Peroxidase GPx3 and Its Role in Cancer. Cancers *12*, 2197. https://doi.org/10.3390/cancers12082197.
67. Bhosale, S.D., Moulder, R., Kouvonen, P., Lahesmaa, R., and Goodlett, D.R. (2017). Mass Spectrometry-Based Serum Proteomics for Biomarker Discovery and Validation. Methods Mol. Biol. *1619*, 451–466. https://doi.org/10.1007/978-1-4939-7057-5_31.
68. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968. https://doi.org/10.1093/bioinformatics/btq054.
69. Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics *30*, 2524–2526. https://doi.org/10.1093/bioinformatics/btu305.
70. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372. https://doi.org/10.1038/nbt.1511.
71. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: A peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794–1805. https://doi.org/10.1021/pr101065j.
72. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics *13*, 2513–2526. https://doi.org/10.1074/mcp.M113.031591.
73. Välikangas, T., Suomi, T., and Elo, L.L. (2018). A systematic evaluation of normalization methods in quantitative label-free proteomics. Brief. Bioinform. *19*, 1–11. https://doi.org/10.1093/bib/bbw095.
74. Suomi, T., Seyednasrollah, F., Jaakkola, M.K., Faux, T., and Elo, L.L. (2017). ROTS: An R package for reproducibility-optimized statistical testing. PLoS Comput. Biol. *13*, e1005562. https://doi.org/10.1371/journal.pcbi.1005562.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides and recombinant proteins		
Acetonitrile (ACN) HPLC gradient grade	SIGMA-ALDRICH	34851
Trifluoroacetic acid (TFA), LC-MS grade	Thermo Scientific	85183
Formic acid, for mass spectrometry ~98%	Fluka	94218-50ML-F
Sequence grade modified Trypsin	Sigma-Aldrich	V5117
DL-dithiothreitol	Sigma-Aldrich	D9779-10G
Iodoacetamide	Sigma-Aldrich	I6125-25G
1 M Tris-HCl, pH 8.0	Invitrogen	15568-025
PEPotec, Grade 2, synthetic peptides with C terminal heavy lysine/arginine	Thermo Fischer Scientific	thermofisher.com/pepotec-srm
Mannitol dehydrogenase	Megazyme (Wicklow, Ireland)	E-MNHPF
1.5 ml Eppendorf tubes	Eppendorf AG	0030 120.086
Sep-Pak® Vax 1cc (50mg) C18 Cartridges	Waters	WAT055955
Critical commercial assays		
Fecal calprotectin ELISA	Calpro AS, Lysaker, Norway	CALP0170
β-Defensin 2 ELISA	Immundiagnostik, Bensheim, Germany	K 6500
Deposited data		
Discovery proteomics data	This study	PRIDE : PXD046387
Targeted proteomics data	This study	https://panoramaweb.org/AG0DKm.url
Software and algorithms		
Excel	2016	https://www.microsoft.com
MaxQuant	1.6.17.0	www.maxquant.net/perseus/
Perseus	1.6.7.0	www.maxquant.net/perseus/
R studio	2022.07.0	www.rstudio.com
Skyline	21.2.0.425 (e653b4c5e)	https://skyline.ms
BioRender	BioRender	https://biorender.com/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Riitta Lahesmaa (rilahes@utu.fi).

Materials availability

This study did not generate new reagents.

Data and code availability

- The proteomics data have been deposited in the PRIDE proteomeXchange repository (https://www.ebi.ac.uk/pride/) and are publically available with the accession numbers PXD046387 and https://panoramaweb.org/AG0DKm.url.
- This paper does not report any original code.
- Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The subject selection, sample collection and dietary randomization were conducted as described in related publications.[21] Briefly, pregnant mothers either with a HLA defined T1D-risk genotype, or with a partner with a HLA defined T1D-risk were recruited. The T1D risk of the offspring was assessed by HLA genotyping from cord blood.

Serum samples were obtained from the mothers at the beginning of the third trimester, at the time of delivery and 3 months postpartum, and from their offspring at birth (cord blood) and at follow-up visits at the ages of 3, 6, 9 and 12 months. The sample series from 22 of these subjects were selected for proteomics analysis, including 12 subjects from the extensively hydrolyzed formula group and 10 from the regular group. In addition, 9-month samples from 40 other subjects in the study were analyzed (14 and 26, regular and extensively hydrolyzed groups, respectively). The formula groups were blinded during selection for LC-MS/MS analysis. The subjects and samples are detailed in **ESM** Tables S4–S6.

This study was conducted with approval of the Ethical Committee of the Pirkanmaa Hospital District (approval number 12/2016). The study was conducted according to the guidelines in the Declaration of Helsinki. Clinical trial registration number: Clinicaltrials.gov Identifier: NCT01735123. Written informed consent was obtained from the families before enrollment. Sex was as defined at birth and all mothers and infants included in the study are Caucasian and white, as that was an inclusion criterion.

Study formula

The intervention formula was an extensively hydrolyzed casein-based formula, while the control formula was composed of 80% intact cow's milk and 20% hydrolyzed milk based formula. The formulas were produced specifically for the study and composed such that the taste and smell would be indistinguishable. The mothers were encouraged to breastfeed as long as possible. The dietary intervention lasted until the infant was 9-months-old and, if the child had not received the study formula for at least 60 days, study formula feeding was continued until 60 days of study formula feeding was reached, but not beyond the age of 8 months. Further details of the dietary data, evaluation of β-cell autoimmunity and intestinal permeability analyses have been described in earlier work.[16]

METHOD DETAILS

Sample preparation

The samples were prepared follows. Serum aliquots (2 μl) were diluted with 8 M urea in 50 mM Tris-HCl pH 8.0, reduced in 10 mM DTT, alkylated with 10 mM IAA in darkness, diluted then digested with trypsin 1:30 ratio (Trypsin:Protein) at 37°C overnight, then quenched with 10% TFA.[67] The digests were desalted using SepPak 50 mg reversed phase SPE cartridges according to established protocols.[67]

Data-dependent LC-MS/MS analysis

Aliquots of the digested serum samples (500 ng) were analyzed by tandem mass spectrometry (LC-MS/MS). An EASY-nLC™ 1200 Ultrahigh pressure liquid chromatograph coupled to a Q Exactive HF quadrapole Oribitrap™ mass spectrometer (Thermo Scientific) was used with a 40 cm x 75 μm i.d. packed capillary column (1.9 μm d_p C_{18} Reprosil particles, Dr Maisch GmbH) with the column heated to 60°C. The peptides were eluted with a gradient from 8 to 20 % B in 18 min, then to 35 % B in 15 min, to 50 % B in 6 min and finally to 100 % in 5.5 min at flow rate of 300 nl/min. The tandem mass spectra were acquired with higher-energy C-trap dissociation (HCD) of the 15 most intense ions (m/z 300-1650, charge states > 1+). The MS1 resolution was set to 120,000, with a 3 x 10^6 AGC target value and a maximum injection time of 55 ms. MS/MS spectra were acquired in the Orbitrap with a resolution of 15,000 (at m/z 200), a target value of 1.2 x 10^3 ions, and a maximum injection time of 120 ms. Dynamic exclusion was set to 20 s. The mobile phase compositions were 0.1% formic acid in MilliQ® water (A) and 0.1% formic acid in 80 % acetonitrile and water (B).

The analyses were carried out using separate batches for each dyad, with randomization and triplicate analysis of each sample. To improve consistency in the protein quantification and allow the transfer of identification between peptide signals, library samples, consisting of control serum depleted of the 12 most abundant proteins (Thermo Scientific™ Pierce™ Top 12 Abundant Protein Depletion Spin Columns) according to the manufacturer's instructions and as previously described,[6] were analyzed with each batch. To avoid the effects of carry-over from establishing cord blood specific proteins, the cord blood samples were analyzed as a separate batch in triplicate. The library samples were also used to establish consistent instrument performance between batches.

Targeted LC-MS/MS analysis

Selected reaction monitoring (SRM) mass spectrometry was used to verify the relative abundance of the targets GGH, PON3, FCN3, PGLYRP2, CERU, MASP1 and the immunoglobulins KVD30, KV240 and LV403, in the serum samples. A1GB, ALB were also measured as reference proteins, and fibrinogen and HBB as quality measures for potential plasma contamination/selection and hemolysis, respectively. Heavy-labeled synthetic peptides (lysine $^{13}C_6$ $^{15}N_2$ and arginine $^{13}C_6$ $^{15}N_4$) peptides were obtained for the targets of interest (*PEPotec,* Grade 2, Thermo Fischer Scientific). The targets were selected on the basis of the detection of their differential expression in the data-dependent LC-MS/MS measurements. Skyline software[68] was used to create the scheduled method for data acquisition and process the data, i.e. to check the peak assignments and their integration.[69]

The trypsin digested samples from both the discovery and follow-up analyses were spiked with synthetic heavy labelled analogues of the peptide targets and a retention time standards (MSRT1, Sigma) for scheduled selected reaction monitoring. The LC-MS/MS analyses were conducted using an Easy-nLC 1000 liquid chromatograph (Thermo Scientific) coupled to a TSQ Vantage Triple Quadrupole Mass Spectrometer (Thermo Scientific). The column configuration included a 20 x 0.1 mm i.d. pre-column in conjunction with a 150 mm x 75 μm i.d. analytical column, both packed with 5 μm Reprosil C_{18}-bonded silica (Dr Maisch GmbH). The following separation gradient was employed at a flow rate of 300 nl/min; from 5% to 21% B in 11 min, then to 36% B in 9 min, to 100% in 2 min, then ending with an 8 min isocratic period. The mobile

phase compositions were the same as those used for the DDA analysis. The estimated injected amounts was 250 ng of endogenous sample, spiked with 50 fmol of synthetic peptides. The raw SRM data are available through Skyline[68] with the dataset identifier https://panoramaweb.org/AG0DKm.url.

Data analysis – LC-MS/MS

The mass spectrometry raw files were processed using MaxQuant software version 1.5.5.1[70] with the Andromeda search algorithm[71] and a UniProt human protein sequence database (May 2019, 19,960 entries plus 245 common contaminants). The search parameters were trypsin digestion with up to two missed cleavages, carbamidomethylation of cysteine as a fixed modification, N-terminal acetylation and methionine oxidation as variable modifications. The peptide and protein level FDR were set to 0.01, based on the use of reversed and true sequence database searches. Label free quantification (MaxLFQ) was used with the match between run options to transfer identifications and calculate protein relative intensity values across the samples.[72]

Protein abundances from technical replicates were averaged and data were normalized using variance stabilization normalization (VSN), which has been shown to perform well in proteomics.[73]

The lactulose–mannitol test

These data are described and reported elsewhere.[56] In brief, after fasting for a minimum of 4 h, the infants were given an oral dose of 2 ml/kg of a lactulose–mannitol solution containing 5 g lactulose and 2 g mannitol per 100 ml. Urine was collected for 5 h, then stored at –20°C. Lactulose concentrations were measured in the samples (25 μl) by the addition of β-galactosidase (12.5 μl), an enzyme cocktail (680 μl), and pepsinogen I (PGI, 20 μl). The working concentrations were of 500 U/ml β-galactosidase, 10 mM ATP, 14.9 mM NADP, 3.64 U/ml HK/G6P-DH, and 350 U/ml for PGI. Absorbance measurements were made with a VICTOR, Wallac 1420 workstation (PerkinElmer, Waltham, MA, United States).

Mannitol concentrations were determined as follows. Sample aliquots (5 ml) were added to the enzyme cocktail (250 ml) and mannitol dehydrogenase (12.5 ml). The final concentrations in the working solution were 6.25 mM NAD+, 6.55 mM ATP, 3.64 U/ml HK/G6P-DH, and 133 U/ml mannitol dehydrogenase.[56] After calculating the proportions of the excreted lactulose and mannitol, the LM ratio was calculated by dividing the lactulose value by the mannitol value.

Analysis of fecal calprotectin and HBD-2

Stool samples from the infants were collected at home and frozen immediately (-20°C). Parents brought the frozen samples to the study center, after which they were then stored at −70°C until analysis. Fecal calprotectin and HBD-2 levels were analyzed from the stool samples with commercial ELISA kits according to manufacturer's instructions (Calpro AS, Lysaker, Norway, and β-Defensin 2 ELISA Kit, Immundiagnostik, Bensheim, Germany).[23,24]

Briefly, approximately 100 mg of feces was obtained from each frozen sample. Extraction buffer was then added at a dilution of 1:50 for both HBD-2 and calprotectin. Fecal material with the extraction buffer was vortexed for 30 seconds and mixing was continued in a shaker at 1000 rpm. for 3 minutes or until solid particles had dissolved. Samples were then centrifuged for 10 min at 10000 g at room temperature and the supernatants were collected and stored at −20°C until measured.

QUANTIFICATION AND STATISTICAL ANALYSIS

Linear Mixed Effects Model

To study formula groups and ageing, and to correct for individual differences between subjects, a linear mixed effects model (LME) was applied for each protein, with the abundance as the dependent variable. Formula, sex, and age were treated as fixed effects and individuals were allowed random intercept. Benjamini-Hochberg adjustment was used to assess false discovery rate.

Correlation Analysis

To determine traits common between the dyads and those carried in the cord blood, Spearman Rank correlation (r_s) was calculated between ranked protein intensities from the different samples, Benjamini-Hochberg adjustment was used to assess false discovery rate.

The influence of formula type on serum proteins

Reproducibility-optimized test statistic (ROTS)[74] was used to compare the protein abundances between formula groups at each time point during the course of sample collection, with additonal focus on a larger subset of samples collected at 9 months. The number of bootstrap resamplings (B) was set to 500 and the largest top list size to be considered (K) was set as 75% of the total number of proteins in the data.

Differences between sample types

The combination of the different sample sources (i.e. mother, child, cord blood) revealed strong contrasts between the proteins detected, hindering direct statistical comparison. To represent proteins specific to the different sample types, the data was analyzed in terms of the proportion of missing values, as shown in Figure 2.