Coeliac disease: what can we learn from prospective studies about disease risk?

*Marisa Stahl, Sibylle Koletzko, Carin Andrén Aronsson, Katri Lindfors, Edwin Liu, Daniel Agardh, on behalf of the TEDDY Study Group**

Paediatric prospective studies of coeliac disease with longitudinal collection of biological samples and clinical data offer a unique perspective on disease risk. This Review highlights the information now available from international paediatric prospective studies on genetic and environmental risk factors for coeliac disease. In addition, recent omics studies have made it possible to study complex interactions between genetic and environmental factors and thereby further our insight into the causes of the disease. In the future, paediatric prospective studies will be able to provide more detailed risk prediction models combining genes, the environment, and biological corroboration from multiomics. Such studies could also contribute to biomarker development and an improved understanding of disease pathogenesis.

Introduction

Coeliac disease is an immune-mediated systemic disorder elicited by the ingestion of gluten-containing grains (wheat, rye, and barley).[1] Coeliac disease is characterised by the presence of tissue transglutaminase autoantibodies (tTGA) and endomysium autoantibodies, signifying an autoimmune process referred to as coeliac disease autoimmunity. Children with coeliac disease autoimmunity can simultaneously have, or develop, characteristic mucosal lesions in the small bowel with shortening of the villi (eg, villous atrophy), compensatory elongation of the crypts, and an increased number of intraepithelial lymphocytes, all of which are indicative of coeliac disease.[2] The variability and progression of these mucosal changes is exemplified in individuals with potential coeliac disease, who have positive tTGA and endomysium autoantibody IgA but typical intestinal architecture, as defined by a Marsh score of 0 or 1.[2,3] The largest prospective 12-year follow-up study of such patients demonstrated that a minority of the individuals progress to villous atrophy, whereas the majority had transient, fluctuating, or even persistent autoantibody positivity without development of villous atrophy.[4] According to the North American Society of Pediatric Gastroenterology, Hepatology, and Nutrition, diagnostic confirmation by intestinal biopsy is the standard of care.[5] However, diagnostic guidelines from the European Society of Paediatric Gastroenterology, Hepatology, and Nutrition give the option to make the diagnosis of coeliac disease without an intestinal biopsy in children with tTGA levels over 10 times the upper limit of normal and positive endomysium autoantibodies on a separate blood draw.[1,2] These highly positive serologies have been shown in European prospective studies to have a high positive predictive value of over 99% for biopsy-confirmed coeliac disease.[2] Although coeliac disease can clinically manifest at any age after gluten introduction, prospective birth cohort studies with longitudinal screening have consistently revealed the highest incidence between age 1 and 4 years for both seroconversion to tTGA positivity (defined as coeliac disease autoimmunity) and confirmed coeliac disease.[6–8] After this early peak, new coeliac disease autoimmunity and coeliac disease cases continue to develop throughout the first decade of life, followed by a decline in subsequent years.[9,10] By contrast, in a large community-based cohort of adults over a 10-year follow-up period, the rate of de-novo seroconversion was 0·3% and confirmed coeliac disease was 0·06%.[11] Additional long-term prospective studies are needed to understand the natural history of seroconversion and confirmed coeliac disease diagnoses during adulthood.[12]

This Review aims to provide a narrative summary of the current evidence on the complex interactions of known environmental and genetic risk factors in coeliac disease development in children and adolescents through the lens of prospective studies. Prospective studies are the ideal platform to study early-life exposures and their

Key messages

- Risk factors and biological markers of coeliac disease autoimmunity and coeliac disease can be studied in prospective birth cohorts.
- 18 paediatric prospective studies have identified several key risk factors and biological markers associated with coeliac disease development during childhood.
- Genetic and demographic factors associated with coeliac disease development in childhood include HLA-DQ haplotypes, non-HLA single-nucleotide polymorphisms, and female sex.
- Environmental factors associated with coeliac disease development in childhood include high or low serum vitamin D levels, early-life infections (particularly enterovirus), a high early-life cumulative gluten intake, and to some degree country of residence.
- The study and use of omic measures in these studies are still somewhat preliminary, but they might provide important biological insight into coeliac disease development.
- Future studies should combine knowledge of the risk factors and protective factors contributing to coeliac disease development to improve risk prediction models and design future coeliac disease prevention trials.

Lancet Child Adolesc Health 2023

Published Online
November 13, 2023
https://doi.org/10.1016/S2352-4642(23)00232-8

*Members are listed in the appendix

Pediatric Gastroenterology, Hepatology, and Nutrition, Digestive Health Institute, Department of Pediatrics, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA (M Stahl MD, Prof E Liu MD); **Department of Pediatrics, Dr von Hauner Kinderspital, LMU University Hospital, LMU Munich, Munich, Germany** (Prof S Koletzko MD); **Department of Pediatrics, Gastroenterology and Nutrition, School of Medicine Collegium Medicum University of Warmia and Mazury, Olsztyn, Poland** (Prof S Koletzko); **Unit of Celiac Disease and Diabetes, Department of Clinical Sciences, Lund University, Malmö, Sweden** (C Andrén Aronsson PhD, Prof D Agardh MD); **Celiac Disease Research Center, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland** (K Lindfors PhD)

Correspondence to:
Assist Prof Marisa Stahl, Pediatric Gastroenterology, Hepatology, and Nutrition, Digestive Health Institute, Department of Pediatrics, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
marisa.stahl@childrenscolorado.org

See Online for appendix

association with disease, along with genetic factors. However, the following points should be noted when considering the informative value of the different studies and interpretation of their results.

First, the most rigorous studies were those designed primarily to explore risk factors for developing coeliac disease and taking into account the number of included cases, known HLA risk, and the length and intensity of follow-up.

Second, many of the known prospective studies were initially designed primarily to evaluate environmental risk factors for other outcomes, such as type 1 diabetes, with coeliac disease as a secondary outcome. Although these are important studies given the scarcity of data on risk factors of coeliac disease, the inclusion criteria and outcome measures might differ from studies primarily designed to examine coeliac disease. Subsequently, certain HLA-DQ risk alleles for coeliac disease could be excluded, potentially reducing the generalisability of the results.

Third, with a peak incidence of coeliac disease autoimmunity occurring at age 1–4 years, potential environmental factors should be assessed from birth onwards, ideally also during pregnancy. Additionally, to correctly determine the temporal relationship between observed risk factors and outcomes, the study design should include prospective, frequent serological surveillance for tTGA, particularly in the first years of life, with timely guideline-directed coeliac disease diagnosis in those who screen positive. Relying solely on case finding based on history or clinical symptoms, or both, without prospective serological surveillance will leave many cases undiagnosed and misclassified as healthy controls. Studies not designed with coeliac disease as a primary outcome could be more prone to

	Location	Design	Inclusion criteria	HLA risk known?*	Years recruiting	Number of participants	Coeliac disease autoimmunity screening?†	Timepoints of screening‡	Coeliac disease diagnosis method§	Coeliac disease protocol?
PreventCD [7,13,16–22]	Croatia, Germany, Hungary, Italy, Israel, Netherlands, Poland, and Spain	Randomised double-blind placebo-controlled trial; early gluten introduction at 4 months of age or standard introduction at 6 months of age	Infants <3 months of age, FDR with coeliac disease	Yes	2007–13	963	Yes	Screened at 4, 8, 12, 18, 24, and 36 months; annually thereafter	Biopsy (except five children diagnosed based on ESPGHAN serologic criteria)	Yes; biopsies reviewed by a local and expert pathologist and given Marsh grading
CeliPrev [8,14,23]	Italy	Randomised controlled trial; early gluten introduction at age 6 months or 12 months	Infants <3 months of age, FDR with coeliac disease	Yes	2003–09	707	Yes	Screened at ages 15 months, and 2, 3, 5, 8, and 10 years	Biopsy (except five children who were symptomatic and improved on gluten-free diet without biopsy)	Yes; biopsies were reviewed by expert pathologist at a single centre with Marsh grading
CDGEMM [15]	USA and Italy	Observational birth cohort study	Infants <6 months of age, FDR with coeliac disease, not yet introduced to solid foods	Yes	2014–22	554	Yes	Screened until age 3 years at intervals of every 6 months, thereafter annually until age 5 years	Biopsy or ESPGHAN serological criteria at discretion of their gastrointestinal clinical provider	Yes; those who screened positive on research assays twice were also screened with an endomysium autoantibody IgA and referred to a paediatric gastroenterologist for clinical evaluation
PROFICEL [24]	Spain	Observational birth cohort study	Infants 4 months of age, FDR for coeliac disease	Yes	2006–10	164	No	NA	No	No
TEF study [25]	Ethiopia	Observational birth cohort study	Infants <6 weeks of age	Yes	2018–22	1046	Yes	Screened at 24, 36, and 48 months	If tTGA positive in two consecutive samples, referred to PCP for further investigation	No

CDGEMM=Celiac Disease Genomic, Environmental, Microbiome, and Metabolomic. ESPGHAN=European Society for Paediatric Gastroenterology, Hepatology, and Nutrition. FDR=first-degree relative. NA=not applicable. PCP=primary care provider. TEF=Traditional Ethiopian Food. tTGA=tissue transglutaminase autoantibody.*Refers to known HLA-DQ genotype risk. †Refers to presence of screening protocol for coeliac disease autoimmunity with autoantibodies. ‡Refers to the included timepoints of autoantibody screening. §ESPGHAN criteria refers to serological criteria for coeliac disease including a tTG IgA >10 times the upper limit of normal and positive antiendomysium IgA on a separate blood sample.

Table 1: **Prospective studies in which coeliac disease is the primary outcome**

these limitations of selection bias, reversed temporality, and misclassification bias.

Finally, results from screening studies with known coeliac disease-permissive HLA-DQ genotypes can be extrapolated to the general population, acknowledging that 30–40% of the general population have a coeliac disease-permissive genotype and the remaining 60–70% are HLA-DQ2 and HLA-DQ8 negative and will not develop coeliac disease, regardless of environmental exposures.

Considering these points, we aimed to review paediatric prospective studies that would be able to fulfil these conditions in the search for potential environmental triggers and predictive biological markers of coeliac disease in early childhood.

Overview of the literature

Of 18 paediatric prospective cohort studies, five had a primary focus on coeliac disease autoimmunity or coeliac disease as an outcome; only four studies aimed to identify environmental triggering factors for coeliac disease: PreventCD,[13] CeliPrev,[14] Celiac Disease Genomic, Environmental, Microbiome, and Metabolomic (CDGEMM),[15] and the Traditional Ethiopian Food study (table 1).[25] PreventCD, CeliPrev, and CDGEMM are the only studies that had a formal coeliac disease diagnosis included within the research protocol. The PROFICEL study investigated the microbiota and environmental factors in a small number of infants with known risk, without systematic screening for coeliac disease autoimmunity or coeliac disease.[24] Another eight studies primarily aimed to identify environmental factors in children with known risk for type 1 diabetes, with development of coeliac disease autoimmunity or coeliac disease as a secondary outcome (table 2). Six of those eight studies were informative regarding our question of environmental risk factors in coeliac disease: The Environmental Determinants of Diabetes in the Young (TEDDY),[26] BabyDiab,[39] Diabetes Autoimmunity in the Young (DAISY),[36] MIDIA (Norwegian acronym for environmental causes of type 1 diabetes),[40] DIABIMMUNE,[46] and type 1 Diabetes Prediction and Prevention (DIPP).[42] Conversely, Celiac Disease Prediction in Skane CiPiS[48,49] and BABYDIET[50] did not routinely screen the whole cohort with tTGA and had a small sample size of children with coeliac disease. Three additional population-based prospective observational birth cohorts that were not designed to examine coeliac disease, but reported factors related to coeliac disease outcomes at specific timepoints, are discussed in this Review but have limitations (table 3). Of those studies, the Generation R study performed screening for coeliac disease autoimmunity in the entire cohort at around age 6 years; those with positive tTGA results were remeasured 3 years later and, if positive results persisted at this timepoint, these patients were further evaluated for coeliac disease.[51] The second cohort, the Norwegian Mother, Father, and Child Cohort Study (MoBa), followed up a nationwide population-based cohort of pregnant women and their offspring; however, this study did not include standardised screening for coeliac disease autoimmunity and, therefore, analyses only included clinically detected cases.[55] The third cohort, All Babies In Southeast Sweden, reported environmental factors in a large prospective birth cohort from the general population, but also did not perform standardised screening for coeliac disease autoimmunity, resulting in a low prevalence of diagnosed coeliac disease in the cohort.[63] Lastly, one randomised controlled intervention trial, the Enquiring About Tolerance (EAT) study, was designed to study the effect of early (age 4–6 months) versus late (after age 6 months) introduction of six different allergenic foods, including wheat, on the development of food allergies.[66] However, tTGA was only tested at age 3 years in this study, which risks missing subsequent coeliac disease autoimmunity cases and incomplete follow-up of those who tested tTGA-positive to determine whether they had confirmed coeliac disease.

Regardless of whether the study was primarily designed to study coeliac disease autoimmunity or coeliac disease, type 1 diabetes, or none of these, very few studies included a coeliac disease diagnosis protocol, leaving the decision to proceed to intestinal biopsy up to clinicians (tables 1–3). Although not all children with coeliac disease autoimmunity develop coeliac disease, coeliac disease autoimmunity can still be used as a proxy to study risk factors that could ultimately lead to coeliac disease. This Review does not include any studies examining risk factors for the development of coeliac disease autoimmunity and coeliac disease during adulthood.

Genetic risk factors

HLA, family history of coeliac disease, female sex, and geographical region are all important factors with variable genetic contributions to the risk of development of coeliac disease.[26] Approximately 50% of the heritability of coeliac disease can be explained by genetic variation, with the majority of the genetic contribution from the HLA-DQ region. Non-MHC loci and additional independent loci within the MHC region, identified through fine mapping and genome-wide association studies (GWAS), also contribute to the heritability of coeliac disease.[67,68] Prospective birth cohort studies enrolling newborns with HLA-DQ2 (which includes combinations of DQ2·5, 2·2, and 7·5) and HLA-DQ8 genotypes have identified children at the highest risk of coeliac disease and allowed risk stratification over 15 years of follow up.[13,14,26] TEDDY enrolled children with coeliac disease-permissive HLA-DQ2 and HLA-DQ8 haplotypes and found that children homozygous for HLA-DQ2 were over 5-times as likely to develop coeliac disease compared with children who were HLA-DQ8 heterozygous or homozygous. HLA-DQ2 heterozygosity conferred an intermediate risk for coeliac disease. This

	Location	Design	Inclusion criteria	HLA risk known?*	Years recruiting	Number of participants	Coeliac disease autoimmunity screening?†	Timepoints of screening‡	Coeliac diagnosis method	Coeliac disease protocol?
TEDDY[4,6,10,26–35]	USA, Germany, Finland, Sweden	Observational birth cohort study	Infants <3 months of age, general population and FDR with type 1 diabetes; HLA screening for type 1 diabetes HLA-DQ	Yes	2004–10	6403	Yes	Every 3 months from age 3 months to 4 years; every 6 months from age 4–15 years	Those with confirmed defined positive tTGA level were referred for clinical evaluation or by self-defined serological criteria	Biopsy (Marsh score ≥2) or high tTGA levels (>100 units for two visits, internally validated)
DAISY[9,16,36–38]	USA	Observational birth cohort study	Infants <3 months of age, general population and FDR with type 1 diabetes; HLA screening for type 1 diabetes HLA-DQ	Yes	1993–2004	1339	Yes	Screened at 9, 15, and 24 months of age, annually thereafter	Those with confirmed positive tTGA were referred for clinical evaluation or by self-defined serological criteria	Biopsy (Marsh score ≥2) or persistently high tTGA levels (tTGA>10 × ULN for two visits)
BabyDiab[39]	Germany	Observational birth cohort study	At birth, FDR with type 1 diabetes	Yes	1989–2000	1511	Yes	Screened at 9 months, 2 years, 5 years, 8 years, 11 years, and 14 years of age	Those with confirmed positive tTGA were referred for clinical evaluation	Biopsy
MIDIA[40,41]	Norway	Observational birth cohort study	Infants from the general population	Yes; only DQ2 or DQ8 genotype enrolled	2001–07	220	Yes	Screened stored samples to find first positive and last negative; collected at 3, 6, 9, and 12 months of age, annually thereafter	Those with confirmed positive tTGA were referred for clinical evaluation	Biopsy or persistently high tTGA levels (tTGA>10 × ULN) and positive HLA-DQ2 or HLA-DQ8
DIPP[42–45]	Finland	Observational birth cohort study	Infants; HLA screening for type 1 diabetes HLA-DQ	Yes	1994–2005	1200	Yes	Annual coeliac disease autoimmunity screening	Those with confirmed positive tTGA were referred for clinical evaluation	Biopsy
DIABIMMUNE[46,47]	Finland, Estonia, Russia	Observational birth cohort study	Either infants or children 3 years of age; HLA screening for type 1 diabetes HLA-DQ	Yes; only DQ2 or DQ8 genotype enrolled	2008–13	5819	Yes	Screened at 3, 6, 12, 18, 24, and 36 months of age	Those with a postive tTGA were referred for clinical evaluation	Biopsy
DiPiS/CiPiS[48,49]	Sweden	Observational birth cohort study	Infants; HLA screening for type 1 diabetes HLA-DQ	Yes; based on B1*02 and B1*03	2001–04	3435	Yes	Screened at 3 years of age with repeat testing if positive	Biopsy (Marsh score ≥1) with improvement in symptoms or serologies after treatment	Biopsy
BabyDiet[50]	Germany	Randomised controlled trial; early gluten introduction at age 6 months or 12 months	Infants <2 months of age, FDR type 1 diabetes; HLA screening for type 1 diabetes HLA-DQ	Yes	2000–06	120	Yes	Screened at 6, 12, 18, 24, 30, and 36 months of age, annually thereafter	Those with confirmed positive tTGA were referred for clinical evaluation	Biopsy

CiPiS=Celiac Disease Prevention in Skane. DAISY=Diabetes Autoimmunity in the Young. DiPiS=Diabetes Prediction in Skane. DIPP=Type 1 diabetes Prediction and Prevention. FDR=first-degree relative. TEDDY=The Environmental Determinants of Diabetes in the Young. tTGA=tissue transglutaminase autoantibody. ULN=upper limit of normal. *Refers to known HLA-DQ genotype risk. †Refers to presence of screening protocol for coeliac disease autoimmunity with autoantibodies. ‡Refers to the included timepoints of autoantibody screening.

Table 2: **Prospective studies in which type 1 diabetes is the primary outcome**

www.thelancet.com/child-adolescent **Published online November 13, 2023 https://doi.org/10.1016/S2352-4642(23)00232-8**

	Location	Design	Inclusion criteria	HLA risk known?*	Years recruiting	Number of participants	Coeliac disease autoimmunity screening?†	Timepoints of screening‡	Coeliac disease diagnosis method	Coeliac disease protocol?
Generation R[51-54]	Netherlands	Observational birth cohort study	Infants	Yes	2002–06	4442	Yes	Screened at 6 years of age with a repeat screen at 9 years of age if positive	Referred for clinical evaluation if tTGA positive at 9 years of age	Biopsy
MoBa[55-62]	Norway	Observational birth cohort study	Pregnant women and their infants	No	2000–09	113 053	No	NA	Clinical case finding	Unknown
ABIS[63-65]	Sweden	Observational birth cohort study	Infants	No	1997–99	16 286	No	NA	Clinical case finding	Biopsy
EAT[66]	UK	Randomised clinical trial; early high-dose gluten introduction at 4–6 months of age compared with after 6 months of age	Infants age 3 months and exclusively breastfed	No	2012–15	1004	Yes	Screened at 3 years of age	Refer to paediatric gastroenterologist if tTGA lever higher than 3 times normal (ie, moderate-to-high positive)	Biopsy or high tTGA levels (tTGA>10×ULN) and positive endomysium autoantibodies IgA in second blood sample

ABIS=All Babies In Southeast Sweden. EAT=Enquiring About Tolerance. MoBa=Mother, Father, and Child Cohort Study. NA=not applicable. tTGA=tissue transglutaminase autoantibody. *Refers to known HLA-DQ genotype risk. †Refers to presence of screening protocol for coeliac disease autoimmunity with autoantibodies. ‡Refers to the included timepoints of autoantibody screening.

Table 3: **Other prospective paediatric studies**

HLA-DQ2 and HLA-DQ8 risk stratification pattern has also been found in other studies, as well as the additional risk conferred for coeliac disease by HLA-DQ7·5 and HLA-DQ2·2 (equivalent to HLA-DQ2·5 in the trans confirmation) and the HLA-DQ2·2 allele more broadly.[7,9,14,69]

However, the heterogeneity in geographical risk for coeliac disease for any given HLA genotype underscores the importance of non-HLA genetic and epigenetic risk factors, as well as environmental factors. Children from different geographical regions with identical HLA-DQ genotypes do not have equivalent coeliac disease risk. In TEDDY, children from Sweden had the highest risk of developing coeliac disease, independent of their HLA-DQ genotype, and children from Colorado had the highest risk in the USA.[10] Other prospective studies have similarly found differences in incidence on the basis of region and ethnicity.[46,52]

GWAS and fine mapping of the MHC region have contributed greatly to the characterisation of the non-HLA-DQ risk factors for coeliac disease and require large numbers of participants, which is not feasible in birth cohort studies.[68,70,71] Over 40 loci outside the HLA region have been shown to be associated with coeliac disease.[72] In TEDDY, 6010 children were analysed using the Immunochip, and 54 single-nucleotide polymorphisms (SNPs) in five genes were associated with coeliac disease development over time (*TAGAP*, *IL18R1*, *RGS21*, *PLEK*, and *CCR9*).[73] SNPs that mapped to the *PKIA* and *PFKFB3* regions met genome-wide statistical significance only in Sweden, suggesting that non-HLA risk might be at least in part accountable for regional variability in coeliac disease incidence. The identified genes have important functions in T-cell activation and homing, IL-2 regulation, and IFNγ production. In other studies using pathway analyses, genetic variation in IFNγ signalling has also been found to be associated with coeliac disease development.[74] Other non-HLA genes have also been implicated in other autoimmune diseases, such as rheumatoid arthritis, type 1 diabetes, and multiple sclerosis, and offer a potential mechanism for co-occurring autoimmune conditions.

Although these analyses have highlighted risk factors for coeliac disease development, other studies have also provided information on potential protective factors. A nested case-control study found that the presence of HLA-DPB1*04:01 could protect genetically susceptible children, specifically those with HLA-DQ2·5, from developing coeliac disease autoimmunity.[75] This finding underscores the importance of understanding the biological interactions of the HLA-DQ region with HLA non-DQ variants for better prospective risk prediction of coeliac disease in genetically susceptible children. Birth cohort studies can also help to validate genetic risk scores by use of genetic data from previous GWAS. The addition of coeliac risk-associated SNPs improves disease prediction compared with HLA risk stratification alone.[16,76] Longitudinal cohort studies in coeliac disease can also serve to identify biological events that occur before disease onset. In PreventCD, circulating microRNAs (miRNAs) were measured in patients with coeliac disease versus control individuals; the study identified 53 statistically significant miRNAs in circulation that represented potential biomarker candidates, including eight miRNAs detected in circulation even before tTGA detection. These findings require further investigation and future implementation strategies to examine how such data could be used for disease prediction in birth cohort studies.[17] In a separate study in an Italian cohort of newborns from families at risk

of coeliac disease and including some children from the PreventCD cohort, expression of nine coeliac disease-candidate genes was analysed at various timepoints: before diagnosis, at the time of coeliac disease diagnosis, and at least 1 year after introduction of a gluten-free diet. In this study, three genes (*KIAA*, *TAGAP*, and *SH2B3*) were found to be overexpressed months before coeliac disease diagnosis. This study had a small patient cohort, with only a single highly variable pre-disease timepoint, and a small number of genes investigated. Nevertheless, this research is an example of how longitudinal birth cohort studies can begin to develop a molecular-based disease prediction model.[18] The further use of omics is discussed later in this Review to highlight the potential of prospective studies.

Environmental risk factors

Prospective studies offer a unique perspective on the complex interplay of genetic and environmental risk factors (figure). The detailed collection of early-life exposures and frequent serial biological testing allow researchers to analyse the relationship between specific risk factors and the development of disease.

Perinatal exposures

Perinatal factors, which include events and conditions occurring during pregnancy and early infancy, have been studied in several registry-based and prospective birth cohorts, with an emphasis on how maternal and infant characteristics and mode of delivery influence the risk of coeliac disease development later in childhood. Maternal factors have been extensively examined without conclusive results. Data linked to medical birth registries have indicated that the risk of coeliac disease in children increases with maternal age, level of education, and maternal coeliac disease.[56,77] However, these findings might also be confounded by factors such as maternal hospitalisation, medication during pregnancy, and mode of delivery, which could not be accounted for in these retrospective studies. In the MoBa cohort, neither maternal infections or maternal antibiotics during pregnancy were associated with coeliac disease in children.[64,57] Smoking during pregnancy also did not affect coeliac disease risk.[58] The mode of delivery has been hypothesised to affect coeliac disease risk due to alterations in the intestinal microbiome. This hypothesis has been extensively investigated in birth registry studies and cohorts and it has been consistently found that caesarean section delivery does not increase coeliac disease risk.[23,28] Thus, there is little evidence from prospective birth cohorts that these perinatal factors might be involved in coeliac disease pathogenesis.

Maternal dietary exposures during pregnancy have been suggested to affect a child's coeliac disease risk. Gluten consumption during pregnancy, particularly in the form of fibre-rich foods and whole-grain cereals, might modify gut-microbiome diversity. So far, evidence is scarce and findings are contradictory. A birth cohort study of children with known genetic coeliac disease risk found no association with maternal gluten consumption, whereas another general population birth cohort study

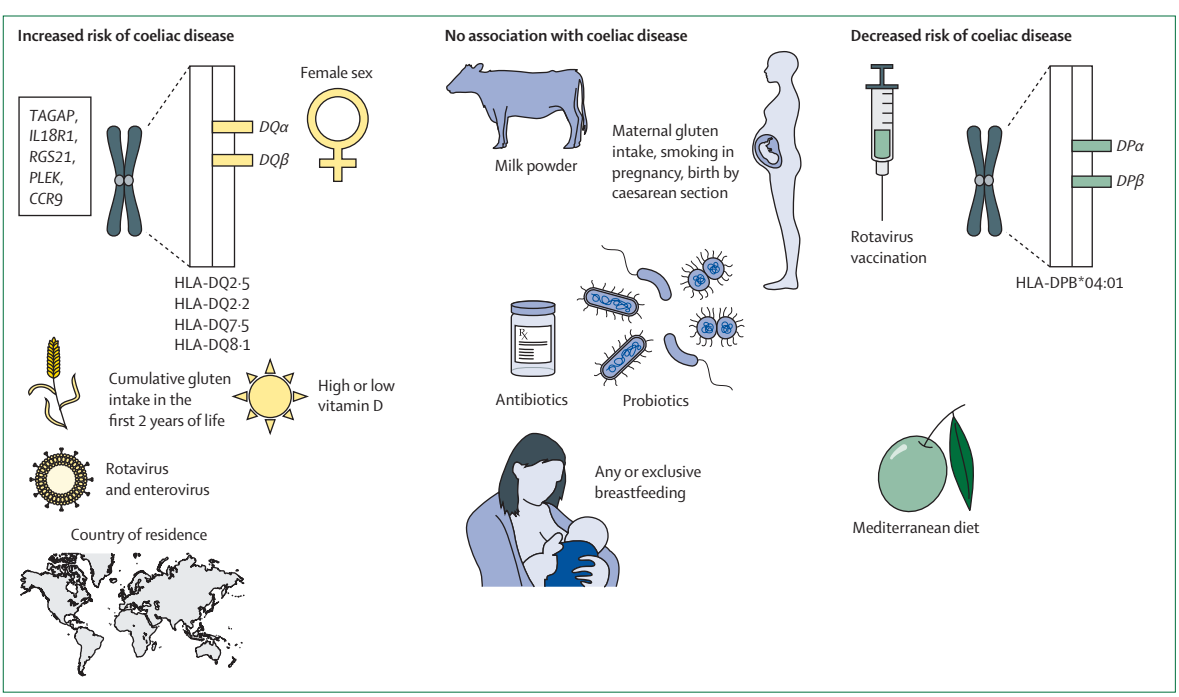

Figure: **Published risk factors from prospective coeliac disease studies affecting later coeliac disease development**
Genetic and environmental risk factors discussed in this Review from prospective coeliac disease studies that increase the risk of coeliac disease (yellow), decrease the risk of coeliac disease (green), or do not affect coeliac disease risk (blue).

found that higher gluten intake during pregnancy was associated with increased coeliac disease risk for children than lower gluten intake during pregnancy.[28,55] Data about maternal dietary supplementation during pregnancy (specifically vitamin D, omega-3 fatty acids, and iron) and risk of coeliac disease in the offspring are also scarce and disparate due to differences in inclusion criteria and study design.[27–29,59,60]

Dietary exposures during early infancy

Human milk contains bioactive and immunological factors that protect against infections and contribute to healthy microbial colonisation, which in turn promotes the growth of beneficial bacteria in an infant's gut. Breastfeeding has been proposed to be potentially protective against coeliac disease development, and the effect of breastfeeding on coeliac disease development has been extensively studied in both randomised controlled trials and cohort studies. These studies have consistently demonstrated that breastfeeding duration (exclusive or any) does not change the risk of developing coeliac disease.[78]

Similar to early introduction of common food allergens to decrease the risk of food allergy, it has been suggested that early gluten introduction could promote tolerance and protect from coeliac disease development.[36] Subsequent observational studies and two randomised controlled trials have shown that timing of gluten introduction (both early and late) is not associated with a cumulative risk of coeliac disease development.[13,14,78] EAT reported that early introduction of high amounts of gluten would decrease the risk of coeliac disease by the age of 3 years.[66] However, this randomised controlled trial was not designed to test coeliac disease as a primary outcome and had major limitations, reducing its generalisability, and caution should be taken in interpretation of the results.

Vitamin D deficiency might also have an important role in coeliac disease pathogenesis in children due to dysregulation of the immune system. Low maternal vitamin D levels during pregnancy, due to low exposure to ultraviolet light, and low or high vitamin D might affect coeliac disease risk in childhood, through an association with the timing of gluten introduction and seasonal viral infections. A nested case-control study of the TEDDY cohort found that both low and high vitamin D serum concentrations were associated with an increased risk of coeliac disease autoimmunity. This finding points to the complexity of the effects of vitamin D on the immune system, at both low and high concentrations, which probably affects coeliac disease risk through immune sequelae.[30]

Dietary exposures during childhood

Dietary patterns can have both inflammatory and anti-inflammatory properties. Western diets, high in red and processed meats, refined grains, simple sugars, and saturated fats, are known to have intestinal inflammatory effects that could in turn be associated with coeliac disease. A Mediterranean-like diet, high in vegetables, grains, and vegetable oils, has been associated with a lower risk of coeliac disease autoimmunity, indicating the importance of studying the whole diet.[53] Consumption of high gluten-containing sources, such as bread and milk-based cereal drinks, during the second year of life was associated with an increased risk of coeliac disease autoimmunity and coeliac disease.[31] In Italian participants of PreventCD, children who developed coeliac disease had dietary intake of higher amounts of refined sugars and lower intakes of legumes, vegetables, and fruits compared with those without coeliac disease.[19]

Although the timing of gluten introduction does not seem to affect coeliac disease risk, a high gluten intake (ie, amounts in the highest compared with the lowest quartile) during the second year of life was consistently associated with both subsequent coeliac disease autoimmunity and coeliac disease development in three prospective studies (TEDDY, PreventCD, and DAISY).[6,19,32,33,37,61,78]

Role of infections

Several studies have investigated the association between infections and coeliac disease, with a particular focus on respiratory and gastrointestinal infections. Many of these studies are limited by a cross-sectional study design and the subsequent conclusions are limited by the risk of reverse causality (ie, infections might actually occur after seroconversion to coeliac disease autoimmunity).[79] Results from prospective birth cohorts with longitudinal testing for coeliac disease autoimmunity have also demonstrated an association between infections and the risk of developing coeliac disease autoimmunity, coeliac disease, or both.[65] However, there has been interstudy variability among birth cohorts regarding the timing and nature of the questionnaires, biological samples used to determine the presence of infection, and the period of follow-up for the study outcomes (appendix pp 3–7). Both PreventCD and MoBa found an association between frequent parent-reported early-life respiratory infections and subsequent coeliac disease.[62,20] By contrast, the TEDDY study found no such association between parent-reported respiratory infection and subsequent coeliac disease autoimmunity.[34]

Both MoBa and TEDDY found an association between parent-reported gastrointestinal infections and subsequent risk of coeliac disease autoimmunity and coeliac disease.[62,34] In TEDDY, rotavirus vaccination was associated with a decreased incidence of coeliac disease autoimmunity. This reduction was only observed in children introduced to gluten before age 6 months, highlighting potential complex interactions between infections and diet in coeliac disease.[34] Similarly, DAISY demonstrated that frequent rotavirus infections determined through the presence of serum rotavirus antibodies were also associated with subsequent coeliac

disease autoimmunity.[38] Varied results from multiple cohort studies and complex interactions with dietary factors suggest that multiple viruses might be implicated and that many environmental risk factors might modify subsequent coeliac disease autoimmunity and coeliac disease development.

In contrast to parent-reported infections only, there are a few prospective study designs that include longitudinal biological samples for more objective evidence of viral infection. Three independent studies (MIDIA, TEDDY, and DIPP) demonstrated an association between enterovirus and subsequent coeliac disease autoimmunity, coeliac disease, or both, through stool samples collected before age 3 years (TEDDY and MIDIA) and serum samples collected 24 months before seroconversion (DIPP).[33,41,47] The Norwegian MIDIA study also found an association between frequent parechovirus infection and subsequent coeliac disease.[40] TEDDY found an interaction between enterovirus infection between age 1–2 years and early-life cumulative gluten intake on subsequent coeliac disease autoimmunity development.[33] Although early studies suggested a possible association with adenovirus serotype 12, this association was not confirmed in two later prospective cohort studies.[41,43]

Some prospective studies have found no association between coeliac disease risk and viral infections, or even a decreased risk with some viral infections on the basis of biological samples. In the Generation R study, cytomegalovirus or combined cytomegalovirus, Epstein-Barr virus, and herpes virus infection was associated with a subsequent decreased incidence of coeliac disease.[54] However, this protective association should be interpreted with caution as Generation R screened for coeliac disease and obtained viral serology at just a single timepoint of age 6 years. DIABIMMUNE analysed nasal swabs and stool samples for common viruses and virus-specific serum antibodies and found no evidence of an association with coeliac disease and viral infection, including both cytomegalovirus and Epstein-Barr virus.[47]

Although mechanistic studies to causally implicate viral infections in coeliac disease pathogenesis are largely absent, and the exact mechanisms by which viruses contribute to the development of autoimmune disease are not fully understood, different pathways have been proposed as potential mechanisms.[80] These pathways include molecular mimicry between viral antigens and self-antigens, viral-mediated destruction of tissue and subsequent release of self-antigens, mucosal barrier dysfunction, and chronic viral infection and shedding, which results in the release of self-antigens (bystander and epitope spreading mechanisms). Moreover, it is plausible that distinct viruses launch a type 1 interferon response, leading to an inflammatory T-cell response and inhibition of regulatory T cells, as has been shown for reovirus-mediated abrogation of gluten tolerance in a murine model.[81]

Thus, several prospective studies indicate that viral infections, particularly enterovirus and rotavirus infections, might be involved in the pathogenesis of coeliac disease. The exact role of viral infections in coeliac disease pathogenesis remains unclear, and there is a paucity of human studies on underlying mechanisms.

Other environmental factors

Other environmental factors during infancy that have been investigated in relation to coeliac disease include season of birth and antibiotic and probiotic exposures. Individuals born in the spring or summer months have a moderately, though statistically significant, increased coeliac disease risk.[82,83] Exposure to some seasonal viral infections during a time window when the infant's immune system develops tolerance to food antigens has been hypothesised to influence coeliac disease risk. This hypothesis is further supported by the interactions observed between season of birth, early infectious exposures, and timing of gluten and coeliac disease autoimmunity development in the TEDDY cohort.[34]

Antibiotics could also disrupt the gut microbiota, leading to dysregulation of the immune system and potentially triggering coeliac disease in genetically susceptible individuals. Although registry-based cohort studies have suggested antibiotic usage as a risk factor for coeliac disease,[77,84,85] a prospective birth cohort could not confirm this; no association was found between cumulative antibiotic use and coeliac disease autoimmunity risk in TEDDY.[34] Similarly, although probiotics might help lead to restoration of the microbiota balance and potentially protect against autoimmunity development, early probiotic use in the first year of life was not associated with a decreased risk of coeliac disease autoimmunity in the TEDDY cohort.[35]

Role of omics in the prediction of coeliac disease

Omics span many different modalities, such as genomics, epigenomics, transcriptomics, proteomics, metabolomics, and microbiome. Omic studies aim to characterise and quantify pools of biological molecules that have roles in the structure, function, and dynamics of a cell, tissue, or organism to study complex biological processes and disease mechanisms holistically. These high-throughput technologies have been used in coeliac disease research, although mostly in cross-sectional studies and certainly not to their full potential in prospective study settings. To best apply omic technologies to coeliac disease prediction, prospective studies that include analysis of samples collected before disease development are essential.

Metabolomics entails the large-scale analysis of small molecules, including amino acids, carbohydrates, nucleotides, lipids, coenzymes, and cofactors within a given compartment. There is a fair number of studies exploiting metabolomics in coeliac disease that vary considerably in terms of the used sample matrix (serum,

plasma, urine, or stool) and analysis methodology; most of these studies were conducted in cross-sectional designs.[86] Only PreventCD and DIPP have used samples collected in prospective birth cohort studies to study metabolomics. PreventCD initially reported no difference in serum metabolites, specifically serum amino acids and polar lipids, between individuals who later developed coeliac disease and those who did not.[21] However, a subsequent subcohort study found lysophostidylcholines, ether-linked phosphatidylcholine, and some phospatidylcholines differed between the two groups.[22] The study did not correct for confounding factors and the small-scale analysis targeted only a narrow spectrum of lipids in different groups. Generally, when interpreting the findings of lipidomic studies, it is important to keep in mind that the differences in the levels of individuals' lipids between the study group and the control group don't allow strong conclusions to be drawn and thus, a broader family-scale analysis of the lipidome would be needed.

DIPP conducted longitudinal lipidomic profiling of plasma samples and reported increased amounts of triacylglycerols and a decreased level of phosphatidylcholines in individuals only 3 months of age, whose diets did not yet include gluten; these individuals later progressed to develop coeliac disease.[44] A further study found that perfluorinated alkyl substances, although not a natural metabolite, were statistically significantly elevated from birth to age 3 months only in the coeliac disease group, and that distinct perfluorinated alkyl substance profiles correlated with triacylglycerol concentrations, particularly of those with saturated fatty acids.[45] Subsequently, exposure to environmental chemicals, such as perfluorinated alkyl substance, might modulate lipid metabolism, which also correlates with later coeliac disease development.

Although the microbiome includes all microorganisms, including bacteria, fungi, archaea, and even viruses in specific anatomical sites, this section will focus on bacteria, as viral infections have been previously discussed. Conducted studies have substantial differences in the applied methodology (high-throughput *vs* low-throughput) and sampling site (duodenum *vs* faeces).[87] Two individual studies focusing solely on the faecal microbiome in paediatric coeliac disease cases followed prospectively from birth have been published. A Finnish study that analysed the stool microbiome from samples collected at age 9 and 12 months revealed no differences between those who later developed coeliac disease and those who did not.[88] By contrast, faecal microbiome profiling of samples collected at age 4 and 6 months, as a part of the prospective PROFICEL study, identified differences for those who subsequently developed coeliac disease.[24] Children who did not develop coeliac disease during the follow-up had an increase in bacterial diversity over time, characterised by increases in particular bacterial strains.

Since the gut microbiome is an independent metabolically active so-called organ with profound effects on host metabolism, it is particularly meaningful to integrate microbiome and metabolomic datasets. Few studies have combined intestinal microbiome characterisation with metabolomics in prospective cohorts. Some of the studies that have combined these modalities were limited either by the number of subjects who developed coeliac disease during the follow-up or by the selected analysis timepoints as some of the individuals had already developed coeliac disease.[89] CDGEMM analysed faecal microbiota, functional pathways, and metabolites in prospectively collected samples, including samples obtained before disease development.[15] The analysis comprised samples collected at the time of diagnosis as well as 3, 6, 9, 12, and 18 months before coeliac disease development. Within-individual analysis in coeliac disease cases revealed an increased abundance of *Dialister invisus, Parabacteroides sp, Lachnospiraceae bacterium*, and serine and threonine metabolites, which have also been linked to other autoimmune and inflammatory conditions and suggest a role of the microbiota and resulting metabolites in coeliac disease development. In the future, multiomic approaches with models that also include environmental exposures and clinical data should also be undertaken. Such multilevel data integration would ensure the most robust model for coeliac disease development. Longitudinal omics studies with larger sample sizes are also needed to validate the referenced findings.

Conclusion

Genetic factors, including HLA-DQ haplotypes, non-HLA SNPs, female sex, and, to some degree, country of residence, play a substantial role in disease risk. Prospective studies have identified environmental factors, including vitamin D levels, early-life infections, specific infant dietary patterns, and, in particular, cumulative gluten intake in the first 2 years of life to be associated with increased risk for coeliac disease development. Other factors, including HLA-DPB*04:01, rotavirus vaccination, and other dietary patterns have been associated with decreased risk. Finally, many other previously reported risk factors have not shown an association in these prospective studies and additional exposomes, including toxic exposures, epigenetics, and biological perinatal exposures, remain unexplored.

The extent to which these environmental factors affect risk in real-world settings and whether they could be exploited to prevent disease remain to be determined. However, these factors will probably need to be accounted for in clinical trials of young children. Of particular interest should be the role of repeated early-life gastrointestinal infections (especially enterovirus) and cumulative gluten intake on the background of genetic risk. The complexities of genetic and environmental interactions supported by biological elements will require machine learning to develop robust disease models of prediction.[90] Herein lies the true potential of prospective cohorts: establishing a

Search strategy and selection criteria

References for this Review were identified through searches of PubMed with the MeSH search terms "coeliac disease" and "cohort" from Jan 1, 1993, until March 4, 2023. This search yielded 2329 articles, from which articles that included prospective studies and risk factors for coeliac disease development were selected. Only papers published in English were reviewed. Articles deemed important for describing potential coeliac disease associations examined later in birth cohort studies, which included national registries and other retrospective data, were also included. Overall, we identified 18 prospective cohort studies that were reviewed in depth.

foundation for risk modelling and biomarker discovery that might also facilitate advances in understanding of disease pathogenesis and development of clinical intervention trials, to prevent coeliac disease in at-risk individuals in the near future.

Contributors

SK, CAA, KL, EL, and DA were involved in the conceptualisation of this Review. All authors were involved in data curation, including the literature search, methodology, and creation of the figure and tables. All authors were involved in writing the original draft and revising of the final version. DA was involved in supervising the writing of the original draft and the revising of the final version.

Declaration of interests

MS has received consulting fees from Pfizer (the Data and Safety Monitoring Board for coeliac disease clinical trials) and Takeda (coeliac disease advisory board). SK has received consulting fees from Danone, has received honoraria from Mead Johnson, Takeda, Nestl.to T Nestlé Nutrition, and Pfizer, and has participated on advisory boards for Takeda, Sanofi, Danone, GlaxoKlineSmith, and Janssen. CAA has grants or contracts with the Swedish Celiac Disease Foundation, Albert Pahlssons Foundation, Maggie Stephens Foundation, Fanny Ekdahls Foundation, Lions Forskningsfond Skane, and Ake Wibergs Foundation. EL and MS have grants or contracts from JDRF and Helmsley Charitable Trust Foundation for the Autoimmunity Screening for Kids study and consult for UptoDate. KL has grants from The Academy of Finland (347473), the Sigrid Juselius Foundation, European Union's Horizon 2020 research and innovation programme (grant agreement No 874864), has received support to attend conferences, serves on the Scientific Advisory Board of The Finnish Celiac Society, and serves as treasurer of the Multi Society Celiac Disease Consortium. EL has received consulting fees from Takeda and serves on the Scientific Advisory Board for the Celiac Disease Foundation and Beyond Celiac. DA has grants or contracts from Lions forskiningsfond Skane, SUS fonder, RoU projektmedel, Swedish Celiac Disease Foundation, Swedish Research Council (Grant 2018-02553; 2022-000537), SFBS Forskningsmedel, ALF projektmedel, Swedish Childhood Diabetes Foundation, EU Interreg, H2020-SCI-2019, NIH/NIDDK 1 R01 DK124581-01, has received support to attend conferences, is co-inventor of a patent with Probi AB, Sweden, has received consulting fees from Takeda and serves on the Scientific Advisory Board for Allero Therapeutics and the Swedish Celiac Disease Research Foundation.

Acknowledgments

The TEDDY Study is funded by U01 DK63829, U01 DK63861, U01 DK63821, U01 DK63865, U01 DK63863, U01 DK63836, U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 DK63821, UC4 DK63865, UC4 DK63863, UC4 DK63836, UC4 DK95300, UC4 DK100238, UC4 DK106955, UC4 DK112243, UC4 DK117483, U01 DK124166, U01 DK128847, and contract number HHSN267200700014C from the National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Allergy and Infectious Diseases, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institute of Environmental Health Sciences, US Centers for Disease Control and Prevention, and JDRF. This work is supported in part by the US National Institutes of Health (NIH)-National Center for Advancing Translational Sciences Clinical and Translational Science Award to the University of Florida (UL1 TR000064) and the University of Colorado (UL1 TR002535). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1 Husby S, Koletzko S, Korponay-Szabó IR, et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 2012; **54:** 136–60.

2 Husby S, Koletzko S, Korponay-Szabó I, et al. European Society Paediatric Gastroenterology, Hepatology and Nutrition Guidelines for Diagnosing Coeliac Disease 2020. *J Pediatr Gastroenterol Nutr* 2020; **70:** 141–56.

3 Ludvigsson JF, Leffler DA, Bai JC, et al. The Oslo definitions for coeliac disease and related terms. *Gut* 2013; **62:** 43–52.

4 Auricchio R, Mandile R, Del Vecchio MR, et al. Progression of celiac disease in children with antibodies against tissue transglutaminase and normal duodenal architecture. *Gastroenterology* 2019; **157:** 413–20.

5 Hill ID, Fasano A, Guandalini S, et al. NASPGHAN clinical report on the diagnosis and treatment of gluten-related disorders. *J Pediatr Gastroenterol Nutr* 2016; **63:** 156–65.

6 Andrén Aronsson C, Lee HS, Hård Af Segerstad EM, et al. Association of gluten intake during the first 5 years of life with incidence of celiac disease autoimmunity and celiac disease among children at increased risk. *JAMA* 2019; **322:** 514–23.

7 Meijer CR, Auricchio R, Putter H, et al. Prediction models for celiac disease development in children from high-risk families: data from the PreventCD cohort. *Gastroenterology* 2022; **163:** 426–36.

8 Lionetti E, Castellaneta S, Francavilla R, Pulvirenti A, Naspi Catassi G, Catassi C. Long-term outcome of potential celiac disease in genetically at-risk children: the prospective CELIPREV cohort study. *J Clin Med* 2019; **8:** 186.

9 Liu E, Dong F, Barón AE, et al. High incidence of celiac disease in a long-term study of adolescents with susceptibility genotypes. *Gastroenterology* 2017; **152:** 1329–36.

10 Stahl M, Li Q, Lynch K, et al. Incidence of pediatric celiac disease varies by region. *Am J Gastroenterol* 2023; **118:** 539–45.

11 Choung RS, Khaleghi S, Cartee AK, et al. Community-based study of celiac disease autoimmunity progression in adults. *Gastroenterology* 2020; **158:** 151–59.

12 Catassi C, Kryszak D, Bhatti B, et al. Natural history of celiac disease autoimmunity in a USA cohort followed since 1974. *Ann Med* 2010; **42:** 530–38.

13 Vriezinga SL, Auricchio R, Bravi E, et al. Randomized feeding intervention in infants at high risk for celiac disease. *N Engl J Med* 2014; **371:** 1304–15.

14 Lionetti E, Castellaneta S, Francavilla R, et al. Introduction of gluten, HLA status, and the risk of celiac disease in children. *N Engl J Med* 2014; **371:** 1295–303.

15 Leonard MM, Valitutti F, Karathia H, et al. Microbiome signatures of progression toward celiac disease onset in at-risk children in a longitudinal prospective cohort study. *Proc Natl Acad Sci USA* 2021; **118:** e2020322118.

16 Romanos J, Rosén A, Kumar V, et al. Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants. *Gut* 2014; **63:** 415–22.

17 Tan IL, Coutinho de Almeida R, Modderman R, et al. Circulating miRNAs as potential biomarkers for celiac disease development. *Front Immunol* 2021; **12:** 734763.

18 Galatola M, Cielo D, Panico C, et al. Presymptomatic diagnosis of celiac disease in predisposed children: the role of gene expression profile. *J Pediatr Gastroenterol Nutr* 2017; **65:** 314–20.

19 Auricchio R, Calabrese I, Galatola M, et al. Gluten consumption and inflammation affect the development of celiac disease in at-risk children. *Sci Rep* 2022; **12:** 5396.

20 Auricchio R, Cielo D, de Falco R, et al. Respiratory infections and the risk of celiac disease. *Pediatrics* 2017; **140:** e20164102.

21 Kirchberg FF, Werkstetter KJ, Uhl O, et al. Investigating the early metabolic fingerprint of celiac disease—a prospective approach. *J Autoimmun* 2016; **72:** 95–101.

22 Auricchio R, Galatola M, Cielo D, et al. A phospholipid profile at 4 months predicts the onset of celiac disease in at-risk infants. *Sci Rep* 2019; **9:** 14303.

23 Lionetti E, Castellaneta S, Francavilla R, Pulvirenti A, Catassi C. Mode of delivery and risk of celiac disease: risk of celiac disease and age at gluten introduction cohort study. *J Pediatr* 2017; **184:** 81–86.

24 Olivares M, Walker AW, Capilla A, et al. Gut microbiota trajectory in early life may predict development of celiac disease. *Microbiome* 2018; **6:** 36.

25 Gudeta AN, Andrén Aronsson C, Binagdie BB, Girma A, Agardh D. Incidence of celiac disease autoimmunity and associations with maternal tuberculosis and pediatric *Helicobacter pylori* infections in 4-year-old Ethiopian children followed up in an HLA genotyped birth cohort. *Front Pediatr* 2022; **10:** 999287.

26 Liu E, Lee HS, Aronsson CA, et al. Risk of pediatric celiac disease according to HLA haplotype and country. *N Engl J Med* 2014; **371:** 42–49.

27 Koletzko S, Lee HS, Beyerlein A, et al. Cesarean section on the risk of celiac disease in the offspring: the Teddy study. *J Pediatr Gastroenterol Nutr* 2018; **66:** 417–24.

28 Uusitalo U, Lee HS, Aronsson CA, et al. Gluten consumption during late pregnancy and risk of celiac disease in the offspring: the TEDDY birth cohort. *Am J Clin Nutr* 2015; **102:** 1216–21.

29 Yang J, Tamura RN, Aronsson CA, et al. Maternal use of dietary supplements during pregnancy is not associated with coeliac disease in the offspring: the environmental determinants of diabetes in the young (TEDDY) study. *Br J Nutr* 2017; **117:** 466–72.

30 Andrén Aronsson C, Liu X, Norris JM, et al. 25(OH)D levels in infancy is associated with celiac disease autoimmunity in at-risk children: a case-control study. *Front Nutr* 2021; **8:** 720041.

31 Hård Af Segerstad EM, Liu X, Uusitalo U, Agardh D, Andrén Aronsson C. Sources of dietary gluten in the first 2 years of life and associations with celiac disease autoimmunity and celiac disease in Swedish genetically predisposed children: the environmental determinants of diabetes in the young (TEDDY) study. *Am J Clin Nutr* 2022; **116:** 394–403.

32 Andrén Aronsson C, Lee HS, Koletzko S, et al. Effects of gluten intake on risk of celiac disease: a case-control study on a Swedish birth cohort. *Clin Gastroenterol Hepatol* 2016; **14:** 403–09.

33 Lindfors K, Lin J, Lee HS, et al. Metagenomics of the faecal virome indicate a cumulative effect of enterovirus and gluten amount on the risk of coeliac disease autoimmunity in genetically at risk children: the TEDDY study. *Gut* 2020; **69:** 1416–22.

34 Kemppainen KM, Lynch KF, Liu E, et al. Factors that increase risk of celiac disease autoimmunity after a gastrointestinal infection in early life. *Clin Gastroenterol Hepatol* 2017; **15:** 694–702.

35 Uusitalo U, Andren Aronsson C, Liu X, et al. Early probiotic supplementation and the risk of celiac disease in children at genetic risk. *Nutrients* 2019; **11:** 1790.

36 Norris JM, Barriga K, Hoffenberg EJ, et al. Risk of celiac disease autoimmunity and timing of gluten introduction in the diet of infants at increased risk of disease. *JAMA* 2005; **293:** 2343–51.

37 Mårild K, Dong F, Lund-Blix NA, et al. Gluten intake and risk of celiac disease: long-term follow-up of an at-risk birth cohort. *Am J Gastroenterol* 2019; **114:** 1307–14.

38 Stene LC, Honeyman MC, Hoffenberg EJ, et al. Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study. *Am J Gastroenterol* 2006; **101:** 2333–40.

39 Hummel S, Hummel M, Banholzer J, et al. Development of autoimmunity to transglutaminase C in children of patients with type 1 diabetes: relationship to islet autoantibodies and infant feeding. *Diabetologia* 2007; **50:** 390–94.

40 Tapia G, Chudá K, Kahrs CR, et al. Parechovirus infection in early childhood and association with subsequent celiac disease. *Am J Gastroenterol* 2021; **116:** 788–95.

41 Kahrs CR, Chuda K, Tapia G, et al. Enterovirus as trigger of coeliac disease: nested case-control study within prospective birth cohort. *BMJ* 2019; **364:** l231.

42 Simell S, Hoppu S, Simell T, et al. Age at development of type 1 diabetes- and celiac disease-associated antibodies and clinical disease in genetically susceptible children observed from birth. *Diabetes Care* 2010; **33:** 774–79.

43 Oikarinen M, Puustinen L, Lehtonen J, et al. Enterovirus infections are associated with the development of celiac disease in a birth cohort study. *Front Immunol* 2020; **11:** 604529.

44 Sen P, Carlsson C, Virtanen SM, et al. Persistent alterations in plasma lipid profiles before introduction of gluten in the diet associated with progression to celiac disease. *Clin Transl Gastroenterol* 2019; **10:** 1–10.

45 Sinisalu L, Sen P, Salihović S, et al. Early-life exposure to perfluorinated alkyl substances modulates lipid metabolism in progression to celiac disease. *Environ Res* 2020; **188:** 109864.

46 Simre K, Uibo O, Peet A, et al. Exploring the risk factors for differences in the cumulative incidence of coeliac disease in two neighboring countries: the prospective DIABIMMUNE study. *Dig Liver Dis* 2016; **48:** 1296–301.

47 Simre K, Uibo O, Peet A, et al. Early-life exposure to common virus infections did not differ between coeliac disease patients and controls. *Acta Paediatr* 2019; **108:** 1709–16.

48 Björck S, Brundin C, Lörinc E, Lynch KF, Agardh D. Screening detects a high proportion of celiac disease in young HLA-genotyped children. *J Pediatr Gastroenterol Nutr* 2010; **50:** 49–53.

49 Björck S, Lynch K, Brundin C, Agardh D. Repeated screening can be restricted to at-genetic-risk birth cohorts. *J Pediatr Gastroenterol Nutr* 2016; **62:** 271–75.

50 Hummel S, Pflüger M, Hummel M, Bonifacio E, Ziegler AG. Primary dietary intervention study to reduce the risk of islet autoimmunity in children at increased risk for type 1 diabetes: the BABYDIET study. *Diabetes Care* 2011; **34:** 1301–05.

51 Jansen M, van Zelm M, Groeneweg M, et al. The identification of celiac disease in asymptomatic children: the Generation R Study. *J Gastroenterol* 2018; **53:** 377–86.

52 Jansen MAE, Beth SA, van den Heuvel D, et al. Ethnic differences in coeliac disease autoimmunity in childhood: the Generation R Study. *Arch Dis Child* 2017; **102:** 529–34.

53 Barroso M, Beth SA, Voortman T, et al. Dietary patterns after the weaning and lactation period are associated with celiac disease autoimmunity in children. *Gastroenterology* 2018; **154:** 2087–96.

54 Jansen MA, van den Heuvel D, van der Zwet KV, et al. Herpesvirus infections and transglutaminase type 2 antibody positivity in childhood: the Generation R study. *J Pediatr Gastroenterol Nutr* 2016; **63:** 423–30.

55 Lund-Blix NA, Tapia G, Mårild K, et al. Maternal fibre and gluten intake during pregnancy and risk of childhood celiac disease: the MoBa study. *Sci Rep* 2020; **10:** 16439.

56 Emilsson L, Magnus MC, Størdal K. Perinatal risk factors for development of celiac disease in children, based on the prospective Norwegian mother and child cohort study. *Clin Gastroenterol Hepatol* 2015; **13:** 921–27.

57 Mårild K, Kahrs CR, Tapia G, Stene LC, Størdal K. Maternal infections, antibiotics, and paracetamol in pregnancy and offspring celiac disease: a cohort study. *J Pediatr Gastroenterol Nutr* 2017; **64:** 730–36.

58 Mårild K, Tapia G, Midttun Ø, et al. Smoking in pregnancy, cord blood cotinine and risk of celiac disease diagnosis in offspring. *Eur J Epidemiol* 2019; **34:** 637–49.

59 Mårild K, Tapia G, Haugen M, et al. Maternal and neonatal vitamin D status, genotype and childhood celiac disease. *PLoS One* 2017; **12:** e0179080.

60 Størdal K, Haugen M, Brantsæter AL, Lundin KE, Stene LC. Association between maternal iron supplementation during pregnancy and risk of celiac disease in children. *Clin Gastroenterol Hepatol* 2014; **12:** 624–31.

61 Lund-Blix NA, Mårild K, Tapia G, Norris JM, Stene LC, Størdal K. Gluten intake in early childhood and risk of celiac disease in childhood: a nationwide cohort study. *Am J Gastroenterol* 2019; **114:** 1299–306.

62 Mårild K, Kahrs CR, Tapia G, Stene LC, Størdal K. Infections and risk of celiac disease in childhood: a prospective nationwide cohort study. *Am J Gastroenterol* 2015; **110:** 1475–84.

63 Ludvigsson JF, Ludvigsson J. Stressful life events, social support and confidence in the pregnant woman and risk of coeliac disease in the offspring. *Scand J Gastroenterol* 2003; **38:** 516–21.

64 Mårild K, Ludvigsson J, Sanz Y, Ludvigsson JF. Antibiotic exposure in pregnancy and risk of coeliac disease in offspring: a cohort study. *BMC Gastroenterol* 2014; **14:** 75.

65 Welander A, Tjernberg AR, Montgomery SM, Ludvigsson J, Ludvigsson JF. Infectious disease and risk of later celiac disease in childhood. *Pediatrics* 2010; **125:** e530–36.

66 Logan K, Perkin MR, Marrs T, et al. Early gluten introduction and celiac disease in the EAT study: a prespecified analysis of the EAT randomized clinical trial. *JAMA Pediatr* 2020; **174:** 1041–47.

67 Kumar V, Wijmenga C, Withoff S. From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. *Semin Immunopathol* 2012; **34:** 567–80.

68 Gutierrez-Achury J, Zhernakova A, Pulit SL, et al. Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. *Nat Genet* 2015; **47:** 577–78.

69 Fernández-Fernández S, Borrell B, Cilleruelo ML, et al. Prevalence of celiac disease in a long-term study of a Spanish at-genetic-risk cohort from the general population. *J Pediatr Gastroenterol Nutr* 2019; **68:** 364–70.

70 Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010; **42:** 295–302.

71 Trynka G, Hunt KA, Bockett NA, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011; **43:** 1193–201.

72 Withoff S, Li Y, Jonkers I, Wijmenga C. Understanding celiac disease by genomics. *Trends Genet* 2016; **32:** 295–308.

73 Sharma A, Liu X, Hadley D, et al. Identification of non-HLA genes associated with celiac disease and country-specific differences in a large, international pediatric cohort. *PLoS One* 2016; **11:** e0152476.

74 Kumar V, Gutierrez-Achury J, Kanduri K, et al. Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels. *Hum Mol Genet* 2015; **24:** 397–409.

75 Hadley D, Hagopian W, Liu E, et al. HLA-DPB1*04:01 protects genetically susceptible children from celiac disease autoimmunity in the TEDDY study. *Am J Gastroenterol* 2015; **110:** 915–20.

76 Sharp SA, Jones SE, Kimmitt RA, et al. A single nucleotide polymorphism genetic risk score to aid diagnosis of coeliac disease: a pilot study in clinical care. *Aliment Pharmacol Ther* 2020; **52:** 1165–73.

77 Canova C, Zabeo V, Pitter G, et al. Association of maternal education, early infections, and antibiotic use with celiac disease: a population-based birth cohort study in northeastern Italy. *Am J Epidemiol* 2014; **180:** 76–85.

78 Szajewska H, Shamir R, Stróżyk A, et al. Systematic review: early feeding practices and the risk of coeliac disease. A 2022 update and revision. *Aliment Pharmacol Ther* 2023; **57:** 8–22.

79 Størdal K, Kahrs C, Tapia G, Agardh D, Kurppa K, Stene LC. Review article: exposure to microbes and risk of coeliac disease. *Aliment Pharmacol Ther* 2021; **53:** 43–62.

80 Smatti MK, Cyprian FS, Nasrallah GK, Al Thani AA, Almishal RO, Yassine HM. Viruses and autoimmunity: a review on the potential interaction and molecular mechanisms. *Viruses* 2019; **11:** 762.

81 Bouziat R, Hinterleitner R, Brown JJ, et al. Reovirus infection triggers inflammatory responses to dietary antigens and development of celiac disease. *Science* 2017; **356:** 44–50.

82 Ivarsson A, Hernell O, Nyström L, Persson LA. Children born in the summer have increased risk for coeliac disease. *J Epidemiol Community Health* 2003; **57:** 36-9.

83 Lebwohl B, Green PH, Murray JA, Ludvigsson JF. Season of birth in a nationwide cohort of coeliac disease patients. *Arch Dis Child* 2013; **98:** 48–51.

84 Mårild K, Ye W, Lebwohl B, et al. Antibiotic exposure and the development of coeliac disease: a nationwide case-control study. *BMC Gastroenterol* 2013; **13:** 109.

85 Dydensborg Sander S, Nybo Andersen AM, Murray JA, Karlstad Ø, Husby S, Størdal K. Association between antibiotics in the first year of life and celiac disease. *Gastroenterology* 2019; **156:** 2217–29.

86 Vacca M, Porrelli A, Calabrese FM, et al. How metabolomics provides novel insights on celiac disease and gluten-free diet: a narrative review. *Front Microbiol* 2022; **13:** 859467.

87 Verdu EF, Schuppan D. Co-factors, microbes, and immunogenetics in celiac disease to guide novel approaches for diagnosis and treatment. *Gastroenterology* 2021; **161:** 1395–411.

88 Rintala A, Riikonen I, Toivonen A, et al. Early fecal microbiota composition in children who later develop celiac disease and associated autoimmunity. *Scand J Gastroenterol* 2018; **53:** 403–09.

89 Sellitto M, Bai G, Serena G, et al. Proof of concept of microbiome–metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants. *PLoS One* 2012; **7:** e33387.

90 MacEachern SJ, Forkert ND. Machine learning for precision medicine. *Genome* 2021; **64:** 416–25.