



Microbiome signatures of progression toward celiac disease onset in at-risk children in a longitudinal prospective cohort study

Maureen M. Leonard^{a,b,c}, Francesco Valitutti^{d,e,1}, Hiren Karathia^{f,1}, Meritxell Pujolassos^{g,1}, Victoria Kenyon^{b,c,1}, Brian Fanelli^{f,1}, Jacopo Troisi^{d,g,h}, Poorani Subramanian^{f,1}, Stephanie Camhi^{b,c}, Angelo Colucci^{g,h}, Gloria Serena^{a,b,c}, Salvatore Cucchiara^l, Chiara Maria Trovatoⁱ, Basilio Malamisura^j, Ruggiero Francavilla^k, Luca Elli^l, Nur A. Hasan^f, Ali R. Zomorodi^{a,b,c}, Rita Colwell^{f,m,2}, Alessio Fasano^{a,b,c,d,2}, and The CD-GEMM Team³

^aDivision of Pediatric Gastroenterology and Nutrition, MassGeneral Hospital for Children, Harvard Medical School, Boston, MA 02114; ^bMucosal Immunology and Biology Research Center, MassGeneral Hospital for Children, Harvard Medical School, Boston, MA 02129; ^cCeliac Research Program, Harvard Medical School, Boston, MA 02114; ^dEuropean Biomedical Research Institute of Salerno, 84125 Salerno, Italy; ^ePediatric Unit, Maternal and Child Health Department, Azienda Ospedaliera Universitaria San Giovanni di Dio e Ruggi d'Aragona, 84125 Salerno, Italy; ^fCosmoID Inc., Rockville, MD 20410; ^gTheoreo srl, University of Salerno, 20851 Salerno, Italy; ^hDepartment of Medicine, Surgery and Dentistry, Scuola Medica Salernitana, University of Salerno, 84090 Salerno, Italy; ⁱPediatric Gastroenterology, Sapienza University of Rome, 00180 Rome, Italy; ^jPediatric Unit, Maternal and Child Health Department, Azienda Ospedaliera Universitaria San Giovanni di Dio e Ruggi d'Aragona, 84125 Salerno, Italy; ^kPediatric Gastroenterology, University of Bari, 84131 Bari, Italy; ^lCenter for Prevention and Diagnosis of Celiac Disease, Fondazione Department and University Hospital (IRCCS) Ca' Granda Ospedale Maggiore Policlinico, 20126 Milan, Italy; and ^mCenter for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742

Contributed by Rita Colwell, April 2, 2021 (sent for review September 28, 2020); reviewed by Philip M. Sherman, Phillip I. Tarr, and Elena F. Verdu

Other than exposure to gluten and genetic compatibility, the gut microbiome has been suggested to be involved in celiac disease (CD) pathogenesis by mediating interactions between gluten/environmental factors and the host immune system. However, to establish disease progression markers, it is essential to assess alterations in the gut microbiota before disease onset. Here, a prospective metagenomic analysis of the gut microbiota of infants at risk of CD was done to track shifts in the microbiota before CD development. We performed cross-sectional and longitudinal analyses of gut microbiota, functional pathways, and metabolites, starting from 18 mo before CD onset, in 10 infants who developed CD and 10 matched nonaffected infants. Cross-sectional analysis at CD onset identified altered abundance of six microbial strains and several metabolites between cases and controls but no change in microbial species or pathway abundance. Conversely, results of longitudinal analysis revealed several microbial species/strains/pathways/metabolites occurring in increased abundance and detected before CD onset. These had previously been linked to autoimmune and inflammatory conditions (e.g., *Dialister invisus*, *Parabacteroides* sp., *Lachnospiraceae*, tryptophan metabolism, and metabolites serine and threonine). Others occurred in decreased abundance before CD onset and are known to have anti-inflammatory effects (e.g., *Streptococcus thermophilus*, *Faecalibacterium prausnitzii*, and *Clostridium clostridioforme*). Additionally, we uncovered previously unreported microbes/pathways/metabolites (e.g., *Porphyromonas* sp., high mannose-type N-glycan biosynthesis, and serine) that point to CD-specific biomarkers. Our study establishes a road map for prospective longitudinal study designs to better understand the role of gut microbiota in disease pathogenesis and therapeutic targets to reestablish tolerance and/or prevent autoimmunity.

gut microbiome | celiac disease | autoimmunity | multiomics analysis

Celiac disease (CD) is a chronic systemic autoimmune disorder that occurs in genetically predisposed individuals and is characterized by loss of tolerance to dietary gluten protein. CD affects ~1% of the global population, with regional variations depending on human leukocyte antigen (HLA) presence and dietary gluten consumption (1). The incidence of CD most likely will continue to increase, along with other autoimmune conditions, despite the fact that its associated genes, HLA, and the trigger, gluten, have not changed (1). Nevertheless, more than 30% of the population carries the predisposing gene and is exposed to gluten, yet only 2 to 3% develop CD in their lifetime (2). This suggests that other factors such as the intestinal microbiota may also contribute to CD pathogenesis.

The inflammatory process underlying CD involves both innate and adaptive immune systems (3). While the adaptive immune response in CD has been described, less is known about the innate immune response following gluten exposure, which drives early steps in CD pathogenesis and eventually leads to loss of gluten tolerance (4). Previous work has linked the trigger of CD, gluten, the intestinal microbiota, and the innate immune response (5–8).

Significance

The incidence of chronic inflammatory autoimmune conditions, such as celiac disease (CD), is increasing at an alarming rate. CD is the only autoimmune condition for which the trigger, gluten, is known. However, its etiology and pathogenesis remain incompletely defined as recent studies suggest other environmental stimuli may play a key role in CD pathogenesis. Here, we prospectively examine the trajectory of the gut microbiota starting 18 mo before CD onset in 10 infants who developed CD and 10 infants who did not. We identified alterations in the gut microbiota, functional pathways, and metabolome before CD onset, suggesting our approach may be used for disease prediction with the ultimate goal of identifying early preventive interventions to reestablish tolerance and prevent autoimmunity.

Author contributions: M.M.L., M.P., J.T., A.C., A.R.Z., and A.F. designed research; F.V., H.K., M.P., V.K., B.F., J.T., P.S., S. Camhi, A.C., G.S., S. Cucchiara, C.M.T., B.M., R.F., L.E., and N.A.H. performed research; M.M.L., B.F., P.S., N.A.H., A.R.Z., R.C., and A.F. analyzed data; T.C.-G.T. contributed to the recruitment and/or data collection/analysis for this project at the international sites; and M.M.L., A.R.Z., R.C., and A.F. wrote the paper.

Reviewers: P.M.S., University of Toronto; P.I.T., Washington University School of Medicine; and E.F.V., McMaster University Medical Centre.

Competing interest statement: M.M.L. serves as a consultant to 9 Meters Biopharma and Anionion and performs sponsored research with Glutenostics LLC. H.K. is a former employee of, B.F. is a current employee of, P.S. is a consultant for, and N.A.H. and R.C. are stockholders at CosmoID Inc. A.F. is a stockholder at Alba Therapeutics, serves as a consultant for Inova Diagnostics and Innovate Biopharmaceuticals, is an advisory board member for Axial Biotherapeutics, and has a speaker agreement with Mead Johnson Nutrition. All other authors have declared that no competing interests exist.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nd/4.0/).

¹F.V., H.K., M.P., V.K., B.F., and P.S. contributed equally to this work.

²To whom correspondence may be addressed. Email: rcolwell@umiacs.umd.edu or afasano@mgh.harvard.edu.

³A complete list of The CD-GEMM Team can be found in *SI Appendix*.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2020322118/-DCSupplemental>.

Published July 12, 2021.

Given the cross-talk between the gut microbiota and immune system, alterations in the gut microbiota have been linked to several autoimmune conditions (9) such as inflammatory bowel disease (10), type 1 diabetes (T1D) (11), multiple sclerosis (12), and CD (13–18). We, and others, have looked for changes in the gut microbiota of infants at risk for CD (15, 17–19). For example, using 16S ribosomal ribonucleic acid (rRNA) amplicon sequencing, we previously reported higher abundance of *Lactobacillus* up to 12 mo of age in one infant who later developed CD compared with 15 at-risk infants who did not (15). Other studies of gut microbiota and CD assessed changes in gut microbiota composition of individuals during the first year after birth who later developed CD compared with controls (17, 18). For example, Olivares et al. (17) used a prospective cohort of 200 infants at risk for CD to compare, with 16S rRNA sequencing, the intestinal microbiota of 10 cases who developed CD during the 5-y study period and 10 matched controls at 4 and 6 mo of age. They reported increases in abundance of *Firmicutes*, *Enterococcaceae*, and *Peptostreptococcaceae* in controls from 4 to 6 mo (17). Rintala et al. (18) also examined intestinal microbiota of infants at risk for CD, using 16S rRNA sequencing, at 9 and 12 mo of age in nine subjects who developed CD by ages 4 and 18 and matched controls, but did not identify any significant differences in microbiota composition. Huang et al. (20) examined intestinal microbiota, using 16S rRNA sequencing, at ages 1, 2.5, and 5 y in 16 subjects with CD (11 of whom developed CD after age 5) and 16 controls and found significant differences in taxonomic composition of the microbiota in cases compared with controls at all of the time points. Finally, a recent study using 16S rRNA sequencing to examine differences in the gut microbiota of children with untreated CD compared with children with treated CD and healthy control subjects did not identify changes in alpha diversity at CD diagnosis (21) but did identify differences in taxonomic composition, such as a lower abundance in *Alistipes* in subjects with CD compared with healthy controls (21). A separate study utilizing 16S rRNA sequencing also identified significant differences in taxonomic composition between patients with newly diagnosed CD and healthy control subjects, with subjects with recently diagnosed CD having a lower abundance of *Bacteroides–Prevotella*, *Akkermansia*, and *Staphylococcaceae* (22).

While these studies provide an important foundation concerning alterations in gut microbiota of subjects at risk for CD early in life, they analyzed only up to three time points in the first year after birth (17, 18, 20) or were restricted to only one subject with CD (15). In addition, the studies used 16S rRNA sequencing to analyze intestinal microbiota, which cannot provide information about functional characteristics of the microbiota nor provide taxonomic data at the strain level, both of which are necessary to design effective treatment for CD. Furthermore, metabolomic analysis (if any) in these studies was generally limited to serum (as opposed to fecal) metabolites, which do not provide direct information about metabolic activity of the gut microbiota. More importantly, here we argue that, to gain mechanistic insight into the pathogenesis of CD and other autoimmune diseases, we need to transition from case–control microbiome studies to prospective longitudinal studies, which prospectively examine subjects at multiple time points before disease development (23). Studies aimed at identifying changes in the microbiome (11, 24) and intestinal permeability (25) have been performed and identified taxonomic changes prior to T1D (11) and necrotizing enterocolitis (24) as well as increased intestinal permeability up to 3 y prior to the development of Crohn’s disease (25). However, longitudinal birth cohort studies focused on multiomic data collection and analysis are limited and have not been developed for CD.

The first step toward achieving this goal was to establish a prospective cohort study for CD, the Celiac Disease Genomic, Environmental, Microbiome, and Metabolomic (CDGEMM) study (26), where we have been following approximately 500 infants in

the United States, Italy, and Spain who have a first-degree relative with CD and therefore, are at a high risk of developing CD. We have previously utilized other study subjects from this cohort and multiomics analysis to investigate the impact of genetic and environmental risk factors on the development of the gut microbiota in infants at risk for CD (19). In the current study, we present proof of concept intersubject and intrasubject analyses using fecal metagenomic and metabolomic data collected at multiple time points before onset of CD in 10 cases and 10 matched controls in order to identify alterations in the intestinal microbiota and metabolome, which may serve as markers of progression toward CD onset.

Results

We included in our study the first 10 children from CDGEMM who had developed CD by the time of inception of the current study (“cases”). We identified matched controls for each of the cases, according to HLA genetics and environmental exposures, to focus on alterations in gut microbiota related to CD. Infants were matched, when possible, by season of birth, location at the time of birth, sex, birth delivery mode, HLA genetics, timing of solid food, and gluten introduction. Characteristics of the 10 cases and 10 matched controls are listed in *SI Appendix, Table S1* and *Dataset S1*. (*Dataset S1* has complete metadata and the algorithm by which cases and controls were matched.)

Fecal samples are collected every 3 mo for the first 3 y after birth and thereafter, every 6 mo until age 5. However, in this study we focused on a time window spanning from 18 mo before time of CD onset until CD onset to ensure that very early changes in the microbiome would be captured. Additionally, the youngest age at which CD was diagnosed in our cohort was 18 mo, further supporting our choice in monitoring microbiome shifts at a very early upstream time point. For the purpose of our study, CD onset is defined as elevated serum antitissue transglutaminase (anti-tTG) and antiendomysial antibody (anti-EMA), measured in our research laboratory, after which subjects are referred for clinical confirmation of CD with additional blood testing and duodenal biopsies as indication following the North American Society for Pediatric Gastroenterology and Nutrition (NASPGHAN) criteria (Hill et al. PMID: 15625418 DOI: 10.1097/00005176-200501000-00001) or the revised European Society for Pediatric Gastroenterology Hepatology and Nutrition (ESPGHAN) criteria (27). For our analysis, we selected time of CD onset as the reference point ($t = 0$) and converted all other sample collection times to relative time points with respect to the time of CD onset. This resulted in six relative time points (in addition to $t = 0$) including -18 , -15 , -12 , -9 , -6 , and -3 mo, with the negative sign implying time before CD onset. A total of 118 fecal samples collected at these time points were analyzed using shotgun sequencing and metabolomic profiling. Taxonomic profiling of the metagenomes was performed at both species- and strain-level resolution for bacteria, archaea, fungi, protists, and viruses. Functional profiling was also done to identify functional (KEGG) pathways encoded by each metagenome, and metabolomics analysis was conducted to profile the metabolites present in each fecal sample (*Methods*). Taxonomic, functional, and metabolite profiling results are available in *SI Appendix, Fig. S1* (*Datasets S2–S4*). The identified taxonomic, functional, and metabolite profiles were extensively analyzed, as described below, to detect and identify intersubject and intrasubject variations in the gut microbiome. Of note, while we did detect human Adenoviruses A, B, C, and D; bacteriophages; fungi; and protists in our samples, we only observed significant changes in one protist species (*Dictyostelium citrinum*) and one virus species (*Human mastadenovirus C*) (*SI Appendix, Fig. S2*) in our analyses. The results presented in the rest of this manuscript are related to bacteria and archaea.

Changes in the Microbiota of Cases Compared with Controls at the Time of CD Onset. We performed a cross-sectional (intersubject) analysis to explore Shannon diversity and Chao 1 richness (*SI Appendix*,

Fig. S3), as well as various features of the gut microbiota (microbes, pathways, and metabolites) differing between cases and controls after a subject develops CD (CD onset) (Fig. 1 and *SI Appendix*, Fig. S2). We did not identify any statistically significant changes in Shannon diversity or Chao 1 richness in cases or controls. The analysis also did not identify microorganisms at the species level that were significantly different in cases compared with controls at CD onset. The analysis did, however, identify a number of microorganisms at the strain level and metabolites whose abundances were significantly different between the two groups (Fig. 1 and *SI Appendix*, Fig. S2). For example, we found that cases had significantly less *Bacteroides vulgatus* str_3775_S_1080 Branch (Fig. 1). A decreased abundance of *B. vulgatus* has been reported to lead to increased gut microbial production of lipopolysaccharide (LPS), which impairs immune function (28). A decreased abundance of *B. vulgatus* has also been reported in infants who developed T1D, compared with matched controls (29). We found that cases also had a significantly decreased abundance of *Bacteroides uniformis* American Type Culture Collection (ATCC)_8492. Previous research in mice has shown that *B. uniformis* decreases tumor necrosis factor- α (TNF- α) production and increases IL-10 production (30), resulting in improved immune defense mechanisms. Thus, a decrease in this strain, as seen in cases, may be associated with a decrease in immune defense mechanisms. Despite several changes at the taxonomic level, we did not identify any functional pathways the abundance of which was significantly different between cases and controls at the time of CD onset. Cross-sectional analysis of metabolites identified several previously unreported metabolites with decreased abundance in cases compared with control subjects at CD onset, including acetyl galactosamine, 2-hydroxyisocaproic acid, and arabinic acid, among others (Fig. 1B). The single identified metabolite implicated in autoimmunity, based on existing literature, was lauric acid, the abundance of which was decreased in cases compared with controls. This observation is in contradiction to previous reports showing that this metabolite in mice is associated with proinflammatory effects by promoting Th1 and Th17 differentiation, resulting in a more severe course of experimental autoimmune encephalitis (31).

Longitudinal Changes in the Microbiota of Cases and Controls. The prospective longitudinal design of our birth cohort study made it possible to perform a longitudinal analysis and gain additional insight beyond a cross-sectional analysis by identifying intrasubject alterations in the gut microbiome before onset of CD. Therefore,

we looked for changes in Shannon diversity and Chao 1 richness, as well as species, strain, pathway, and metabolite changes in abundance differentially between a preonset time point (i.e., -18, -15, -12, -9, -6, and -3 mo) and CD onset (i.e., $t = 0$). This analysis was done for cases and controls separately, and we report only nonoverlapping longitudinal patterns between cases and controls (Figs. 2–5). While we observed statistically significant changes at a number of time points in Shannon diversity and Chao 1 richness for species and strains in both cases and controls, the amount of change in diversity was negligible (*SI Appendix*, Fig. S3). From longitudinal analysis of microbial species (Fig. 2), we found increased abundance of a number of species previously associated with other autoimmune conditions in cases compared with CD onset, which suggests these microorganisms can serve as biomarker of future autoimmune disease. For example, our analysis identified a significantly higher abundance of *Dialister invisus* strain DSM_15470 in CD at all time points (except -3 mo) compared with CD onset. An increased abundance of *Dialister* has been previously reported in children with pre-T1D compared with controls (32) and in subjects who later developed CD (20). We also observed increased abundance of *Parabacteroides* species and strains prior to CD onset in cases (Figs. 2 and 3). Specifically, in the longitudinal analysis, an increase in *Parabacteroides* sp. was noted at all of the time points, except -18 mo, compared with CD onset. An increased abundance of *Parabacteroides* has previously been linked to autoimmune conditions such as T1D (11) and Behcet's disease (33). However, *Parabacteroides distasonis* showed decreased abundance at all preonset time points except -15 mo. Finally, *Lachnospiraceae bacterium* presented increased abundance at all time points except -15 and -6 mo. *L. bacterium* colonization has been associated with obesity and diabetes in genetically at-risk mice (34) and has also been shown to induce colonic inflammation by recruiting macrophages into the colon in the presence of colonic epithelial cell disruption (35).

Our longitudinal analysis also detected a number of species and strains previously identified as having anti-inflammatory properties, whose abundance was lower during the “march” from preclinical to CD onset (Figs. 2 and 3 and *SI Appendix*, Fig. S2). For example, there was a decreased abundance of *Streptococcus thermophilus* at -18, -12, and -6 mo. However, the abundance was higher at all of the other time points. *S. thermophilus* has been identified as a probiotic, releasing an anti-inflammatory metabolite capable of crossing the intestinal barrier (36). In addition, *Faecalibacterium prausnitzii* showed lower abundance at -15 and -12 mo compared

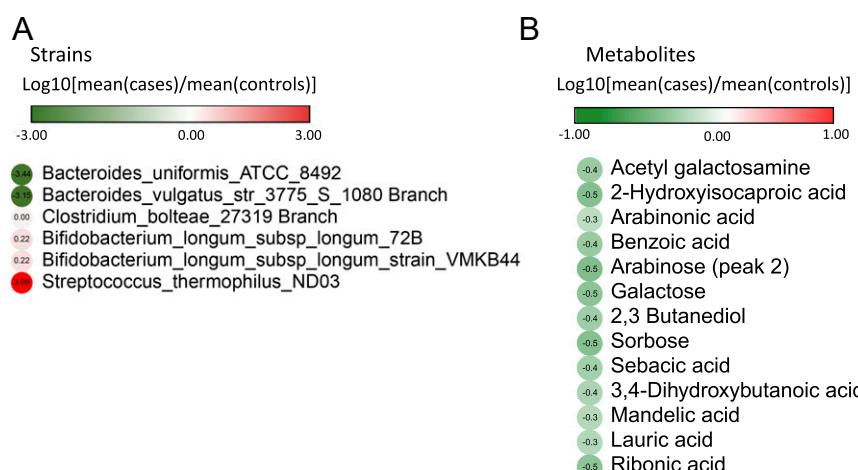


Fig. 1. Cross-sectional (intersubject) analysis of microbiota features at CD onset. Cross-sectional analysis comparing cases and controls at CD onset was performed by using the Mann-Whitney U (Wilcoxon rank-sum) test, and significant results are reported for (A) microbial strains (P value < 0.05) and (B) metabolites (P value < 0.05). No significant species or pathways (P value < 0.05) were detected. Box plots for significant features are shown in *SI Appendix*, Fig. S2.

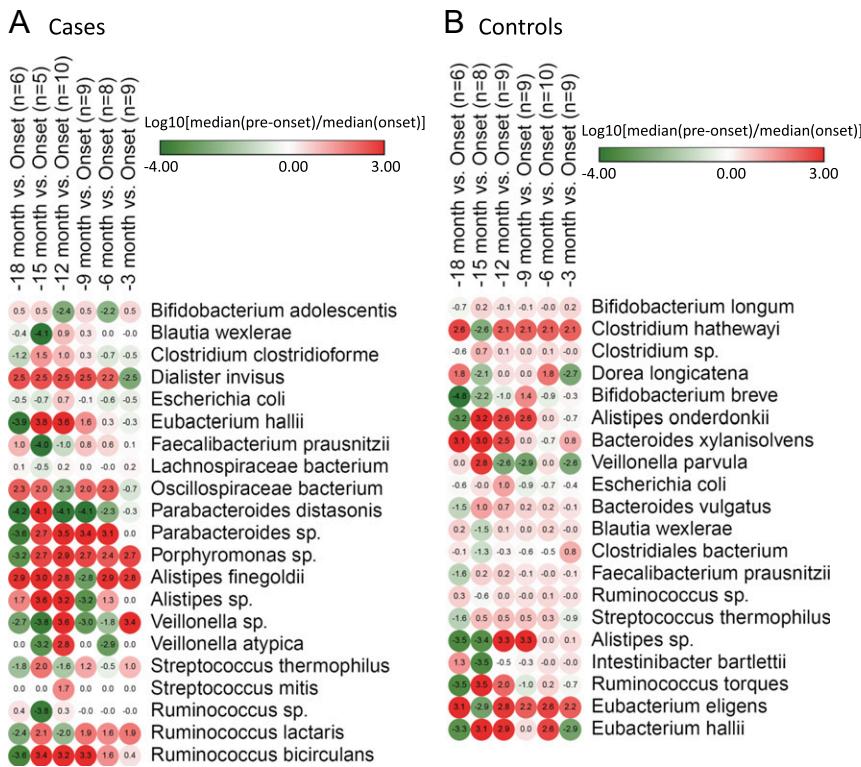


Fig. 2. Longitudinal (intrasubject) analysis for microbial species. A paired Wilcoxon (Wilcoxon signed rank) test was used to identify microbial species whose abundance differentially changes between a preonset time point (-18 , -15 , -12 , -9 , -6 , and -3 mo) and CD onset. Any species for which a statistically significant (P value < 0.05) change is observed in at least one time point in (A) cases and (B) controls is reported here. Box plots for significant features are shown in *SI Appendix, Fig. S2*. Time points at which a significant change is observed are shown in *SI Appendix, Fig. S5*. Here, we report only species for which significant changes are uniquely observed in cases or in controls.

with CD onset, although higher abundance at all other time points. *F. prausnitzii* is known to have anti-inflammatory properties by the release of metabolites capable of blocking nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) activation and interleukin (IL)-8 production (37, 38) and has been reported to be underabundant in fecal samples of subjects with inflammatory bowel disease (IBD) (39–41). An increased abundance of *F. prausnitzii* has been reported in pediatric subjects with newly diagnosed Crohn's disease, which is in agreement with our findings in the year prior to CD onset; however, these findings were noted in the ileal mucosa of patients with Crohn's disease, and therefore, findings are not comparable (41). We also found *Clostridium clostridioforme*, a microorganism that contributes to butyrate production (42), showing decreased abundance at all time points except -15 and -12 mo. Decreased abundance of *C. clostridioforme* in subjects with IBD has been previously reported (43). Previously unreported species and strains were detected in this study, including *Ruminococcus lactaris* (-18 and -12 mo), *Blautia wexlerae* (-15 mo), and *Alistipes finegoldii* (-9 mo) among others, to be significantly reduced in abundance compared with CD onset (Figs. 2 and 3). However, a decrease in abundance of *Alistipes* has been previously reported in a case-control study of subjects with CD compared with control subjects (21).

Longitudinal analysis of microbiome-encoded functional metabolic pathways revealed several that were altered in cases prior to CD onset (Fig. 4 and *SI Appendix, Fig. S2*). Notably, the majority (e.g., tryptophan, LPS biosynthesis, and fatty acid metabolism), statistically significant compared with CD onset, was at -15 mo. Glycophospholipid biosynthesis globo-series and high mannose-type *N*-glycan biosynthesis-related pathways increased at both -15 and -6 mo compared with CD onset. *N*-glycan has been linked to T cell-mediated autoimmune diseases (44), such as T1D (45),

IBD (46), and multiple sclerosis (47), via mechanisms such as increasing the production of interferon gamma (IFN γ) by T cells (48) and decreasing cluster of differentiation 152 (CTLA-4) surface tension (49). Decreased abundance of ether lipid metabolism pathway at -15 and -3 mo was noted. Previous work has shown decreased abundance of ether lipids in serum of children with T1D compared with healthy controls (50), although the association between microbial pathways for ether lipid metabolism in the gut and serum levels of ether remains to be determined. Other pathways not previously reported include glutathione metabolism, phenylalanine metabolism, and tyrosine metabolism, among others.

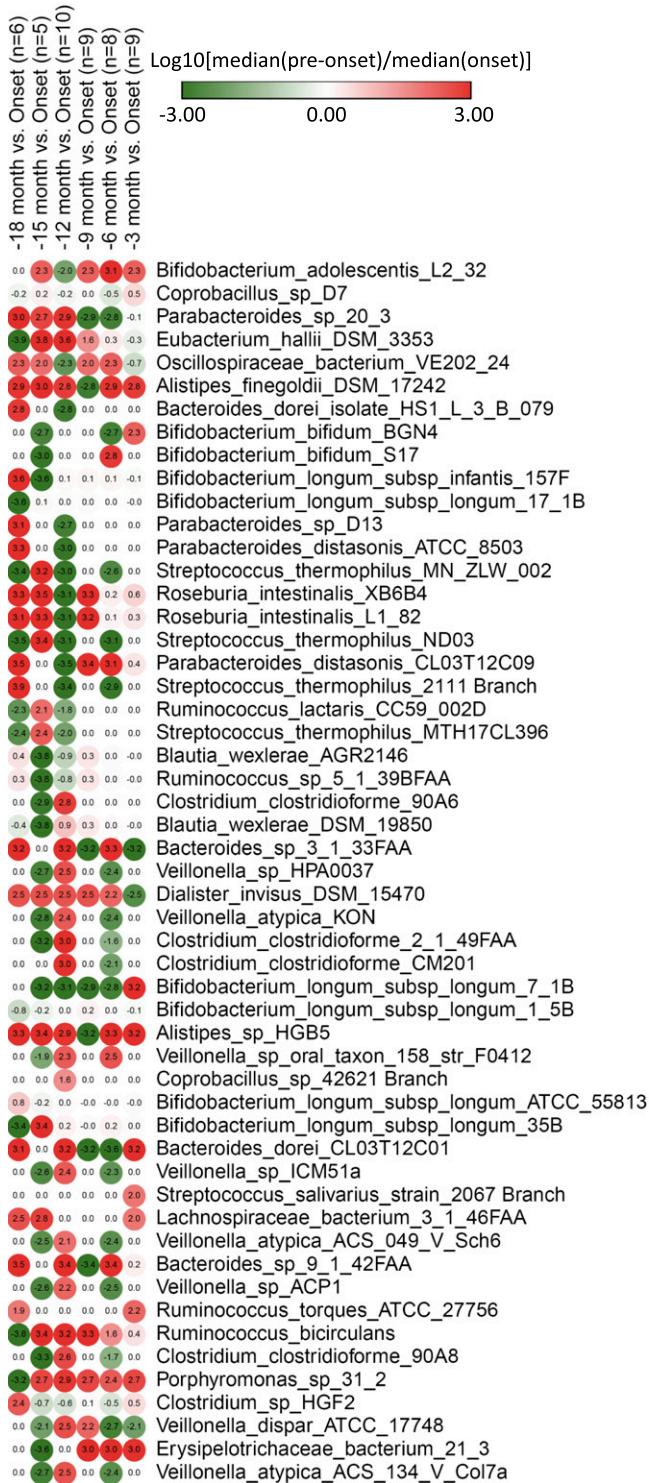
Longitudinal analysis of metabolites identified four metabolites altered in CD cases, including glycolic acid, which was decreased at all time points compared with CD onset, as well as serine, threonine, and 3-hydroxyphenylacetic acid, which were increased at -12 and -15 mo compared with CD onset (Fig. 5A and *SI Appendix, Fig. S2*). Extracellular serine regulates the adaptive immune response because of its essential role in stimulating effector T cell expansion (51). Decreased threonine has previously been reported in serum of patients with rheumatoid arthritis, but it is unclear how findings related to altered serum metabolites compare with altered fecal metabolites (52). In addition, pretreatment of peripheral blood mononuclear cells (PBMCs) with microorganism-derived 3-hydroxyphenylacetic acid has been shown to reduce inflammatory cytokine production following PBMC stimulation by LPS, suggesting this metabolite may have an anti-inflammatory role (53).

Our longitudinal analysis also revealed that species and strains such as *Bifidobacterium longum*, *Bacteroides xylanisolvans*, and *Clostridiales* increased in abundance at -3 mo and other time points compared with $t = 0$ control subjects (CD onset in matched cases).

In particular, six strains of *B. longum* were detected at significantly increased abundance in healthy controls at -3 mo. These findings are in agreement with previously published reports showing

increased abundance of *B. longum* in control subjects compared with those later developing CD (17). *B. longum* has been shown to increase production of IL-10 and decrease inflammatory cytokines

A Cases



B Controls

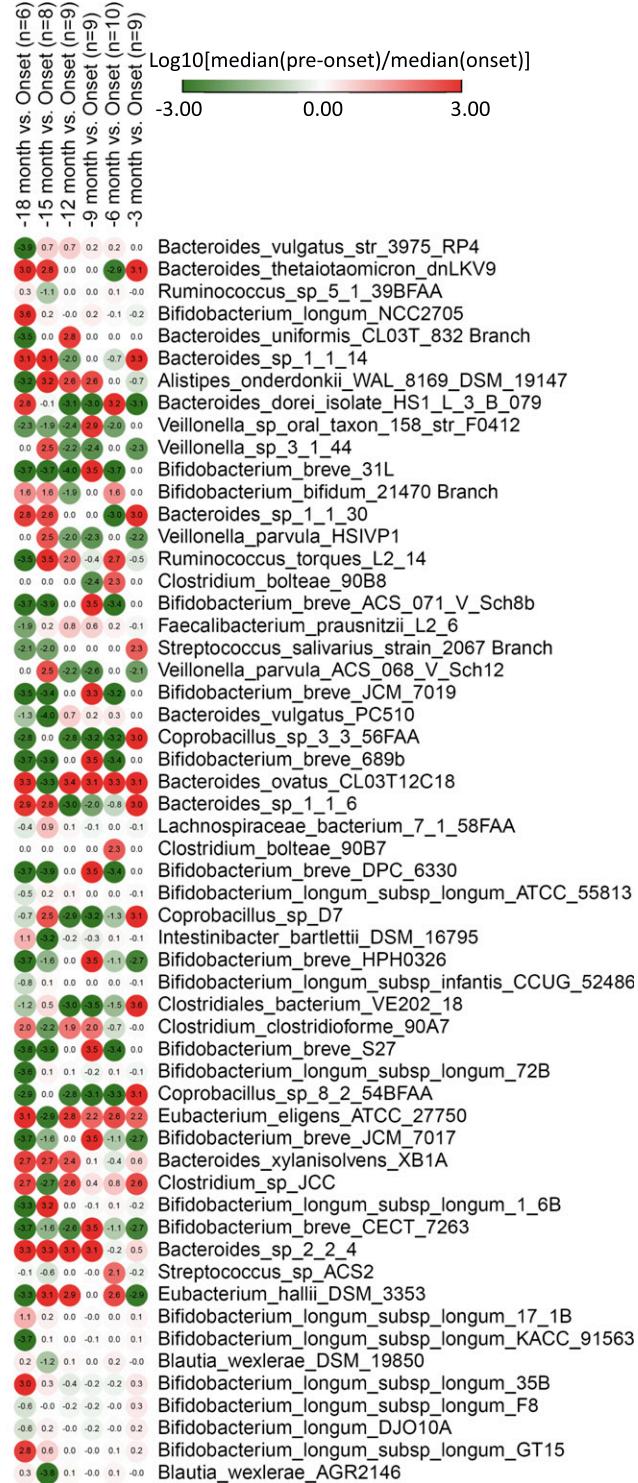


Fig. 3. Longitudinal (intrasubject) analysis for microbial strains. Microbial strains with a statistically significant (Wilcoxon signed rank test, P value < 0.05) change in abundance in at least one preonset time point compared with CD onset in (A) cases and (B) controls. Box plots for significant features are shown in *SI Appendix*, Fig. S2. Time points at which a significant change is observed are shown in *SI Appendix*, Fig. S5. Only strains with a uniquely observed change in cases or controls are reported.

and the CD4+ T cell immune response in animal models with gliadin-induced enteropathy (54, 55), further supporting its possible role in protection against chronic immune conditions. Increased

abundance of *Bifidobacterium breve* and nine strains of the species was detected at -9 mo. *B. breve* is a common probiotic for infants and has been linked to protection against necrotizing enterocolitis

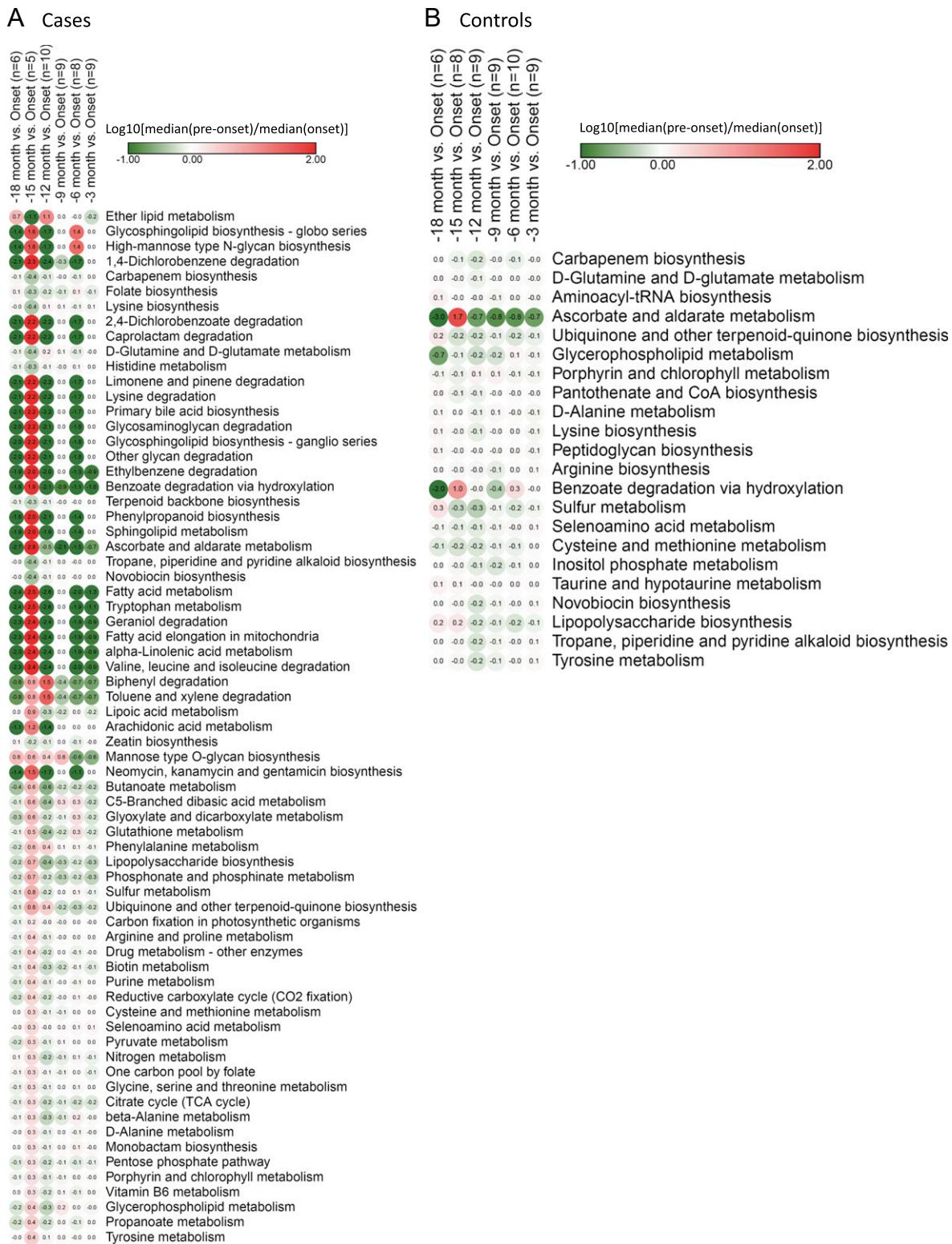


Fig. 4. Longitudinal (intrasubject) analysis for microbial pathways. Microbial pathways with a statistically significant (Wilcoxon signed rank test, P value < 0.05) change in abundance in at least one preonset time point compared with CD onset in (A) cases and (B) controls. Box plots for significant features are shown in *SI Appendix, Fig. S2*. Time points at which a significant change is observed are shown in *SI Appendix, Fig. S5*. Only pathways with a uniquely observed change in cases or controls are reported.

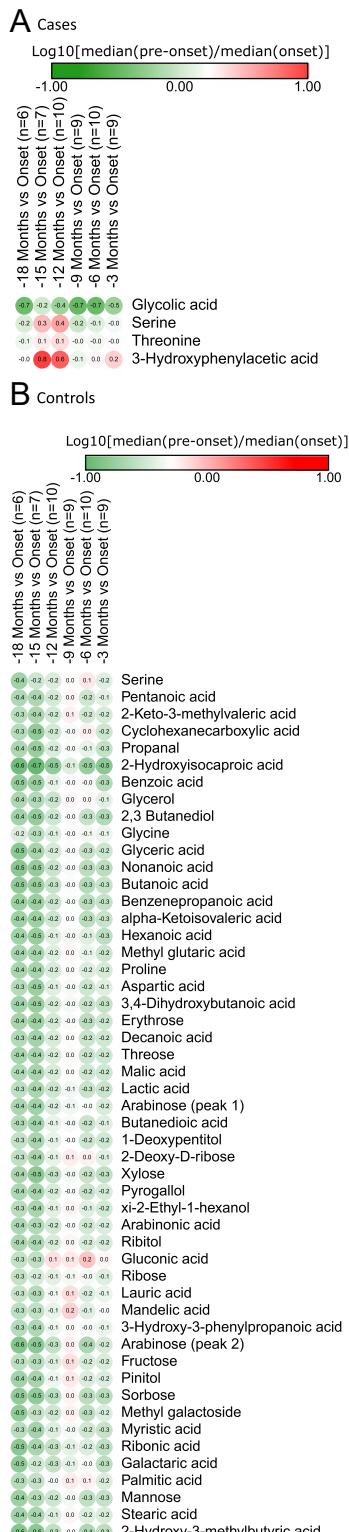


Fig. 5. Longitudinal (intrasubject) analysis for metabolites. Metabolites with a statistically significant (Wilcoxon signed rank test, P value < 0.05) change in abundance in at least one preonset time point compared with CD onset in (A) cases and (B) controls. Box plots for significant features are shown in *SI Appendix, Fig. S2*. Time points at which a significant change is observed are shown in *SI Appendix, Fig. S5*. Similarly, only metabolites with a uniquely observed change in cases or controls are reported.

and development of allergic diseases (56), as well as decreased inflammation in CD (57, 58). Control subjects in our study had a significantly increased abundance of *Escherichia coli* at -12 mo compared with $t = 0$, even though abundance at all other time points was decreased. *E. coli* has been shown to stimulate B regulatory cells to produce anti-inflammatory cytokines and promote development of T regulatory cells, among other mechanisms, to mitigate the inflammatory response (59).

Increased abundance of *Clostridium hathewayi* and *Eubacterium eligens* at all of the points in time was noted in controls compared with $t = 0$ except -15 mo. *E. eligens* promotes production of anti-inflammatory cytokines in vitro (60). *Veillonella parvula*, a microorganism associated with relapsing polychondritis [an autoimmune condition (61)], significantly decreased in abundance at -12 , -9 , and -3 mo compared with $t = 0$.

In contrast to our findings of a significant change in most pathways at -15 mo, longitudinal analysis of microbiome-associated functional pathways in control subjects showed a statistically significant change in abundance of pathways occurring at -3 mo, but the observed fold change was minimal. Only benzoate degradation via hydroxylation and ascorbate and aldarate metabolism, at -15 mo, notably increased in abundance. However, ascorbate and aldarate metabolism was decreased at all other time points compared with $t = 0$. Utilizing ribonucleic acid sequencing (RNAseq), we reported earlier that ascorbate and aldarate metabolism is down-regulated in the intestinal mucosa of subjects with active CD, compared with subjects in remission (5). Finally, our analysis indicated decreased sulfur metabolism (at all time points except -18 mo) and LPS biosynthesis (at all time points except -18 and -15 mo), with increase associated with T1D (62) and autoimmune hepatitis (63), respectively.

Our longitudinal metabolomic analysis of control subjects identified 2-hydroxy-isocaprylic acid, the abundance of which decreased at all time points, compared with CD onset. Other metabolites were detected showing decreased abundance at -18 , -12 , -15 , and -3 mo and slightly increased at -9 and -6 mo compared with $t = 0$ (Fig. 5).

Linking Microbial Species, Pathways, and Metabolites. Connections between bacterial species, pathways, and/or metabolites showing altered abundance before CD onset were determined by performing an association study (using Spearman correlation coefficient) between species/strains and pathways or metabolites (*Datasets S5* and *S6* and *SI Appendix, SI Text* have details). For example, positive association at -12 mo was determined for *Bifidobacterium adolescentis* and high mannose-type *N*-glycan biosynthesis, both increasing in abundance at several time points before CD onset (Figs. 2 and 4). As noted above, *N*-glycan has been linked with autoimmune conditions (44–47), whereas *B. adolescentis* has been reported to increase in subjects with CD (64), implying association with increased risk of developing autoimmunity and thereby, CD because it either contributes directly to high mannose-type *N*-glycan biosynthesis in the gut or positively interacts with other microorganisms responsible for this pathway. Association analysis of microorganism and metabolite also identified a positive association of *L. bacterium* with serine at all time points from -15 to -3 mo in cases. We observed increased abundance of *Lachnospiraceae* species in cases at all time points except -6 and -15 mo. *Lachnospiraceae* and serine have been linked to inflammatory conditions (34) and regulation of the adaptive immune response (51), respectively. This suggests that abundant presence of *Lachnospiraceae* increases risk of developing autoimmune and inflammatory conditions such as CD by either producing serine or positive interaction with gut microbes to produce serine. Other interesting examples include positive association between *B. longum* (specifically subspecies *longum* ATCC 55813), increased in controls at all time points except -18 mo, and *N*-acetyl-D-galactosamine, the latter having been reported to inhibit expression

of the proinflammatory cytokine TNF- α (65). We also observed associations between some metabolites and functional pathways. For example, we identified a positive association between 3-hydroxyphenylacetic acid and the “ubiquinone and other terpenoid-quinone biosynthesis” pathway. Both the metabolite and the pathway were found to be decreased in cases at -3, -6, and -9 mo before CD onset. Of note, ubiquinone has been reported to be an antioxidant (66) and to modulate the innate immune response (67).

Identifying Interactions between Gut Microbiota and Human Host. Pathway enrichment analysis was done to identify human metabolic pathways by which metabolites with significantly altered abundance, determined by cross-sectional and longitudinal analysis, are enriched. The analysis did not identify pathways for metabolites from the cross-sectional analysis. However, it did identify three human pathways, namely fatty acid biosynthesis, pentose phosphate pathway, and aminoacyl-transfer ribonucleic acid (tRNA) biosynthesis, in which significant metabolites from the longitudinal analysis were found to be enriched (P value < 0.05) (SI Appendix, Fig. S4). This result implies that human pathways are affected by microbial metabolites. Interestingly, all three pathways are associated with inflammation. Fatty acid biosynthesis is involved in generating inflammatory mediators, including prostaglandins and cytokines (68). The pentose phosphate pathway regulates immune response by affecting oxidative stress and has been linked to autoimmune conditions, namely rheumatic diseases (69). Finally, aminoacyl-tRNA biosynthesis has been reported to be involved in autoimmune diseases (70), as well as inflammatory processes (71).

Discussion

Most of the literature on the microbiome and the role of intestinal microbiota in disease reports case-control studies aimed at identifying alterations in the intestinal microbiome that can be linked to a specific disease. However, whether such changes are a cause or consequence of disease remains unknown. Here, we were able to include a set of analyses in an ongoing study of a prospective birth cohort of infants at risk of developing CD, allowing a proof of concept metagenomic and metabolomic longitudinal study. It provided an opportunity to observe alterations in the gut microbiota in the earliest phases of CD development because each subject serves as its own control. Recently a similar approach has been taken to study gut microbiota in infants at risk for T1D (72) in the first 5 y after birth. However, T1D is a complex chronic immune condition that peaks between 10 and 14 y of age and is suspected to have an environmental trigger yet to be determined (73). CD, in contrast, is the only autoimmune disorder for which the trigger is known. Furthermore, previous prospective birth cohort studies that did not examine the microbiome but did monitor at-risk infants over a 10-y period found that 80% of infants who developed CD did so by 36 mo of age (74, 75). This is unlike other autoimmune conditions such as T1D (73), IBD (76), or multiple sclerosis (77), for which typical onset is adolescence or adulthood. These findings suggest that CD can serve as an ideal model of chronic immune-based disorders studied prospectively to identify key steps in the progression to the disease, including microbiota surveillance, involving loss of tolerance to gluten.

While a number of studies have begun to evaluate the role of gut microbiota in CD prospectively, either they have been limited to the first year after birth, when the disease actually emerges years later (17, 18), or they include a limited, sparse set of samples collected at specific time points (20) or a single subject (15). Other recent studies describe alterations in the gut microbiota at CD diagnosis compared with in subjects with treated CD and/or healthy controls (21, 22). However, given the case-control design, the studies' focus is on the time of CD diagnosis, which may be different than the time of CD onset since they were not prospectively monitored. Furthermore, the studies did not integrate fecal microbiota analysis with untargeted fecal metabolomic profiles

(17, 18, 20) and used 16S rRNA sequencing for microbiome profiling, which is not capable of capturing strain-level differences and functional characterization relative to the gut microbiota, both essential for designing effective microbiome-based treatment.

The research presented here comprised cross-sectional analysis of cases and comparison of matched controls at time of CD onset. Cases were found to carry a lower abundance of strains of many microbial species, notably *B. vulgaris* _str_3775_S_1080 branch and *B. uniformis* _ATCC_8492, for which decreased abundance has been reported to be associated with impaired immune function (28, 30). However, cross-sectional analysis did not identify changes in species or pathway abundances. Conversely, longitudinal analysis of microbial species, pathways, and metabolites, employing the same statistical methods, revealed major shifts in gut microbiota profiles and associated metabolomes before onset of CD. These findings indicate that longitudinal study is preferable for identifying significant microbial species/strain shifts and relationships, with respect to microbiome composition and function linked to break in gluten tolerance and onset of CD. For example, some microbial species, pathways, and metabolites were found to increase in abundance before the onset of CD, and these had previously been linked to inflammatory and autoimmune conditions. We also identified others with decreased abundance prior to CD onset previously reported to have anti-inflammatory properties. In addition, several other species/pathways/metabolites were detected, which have not been reported before and are presumably CD specific. Examples are increased abundance of *D. invisus*, *Parabacteroides* sp., and *L. bacterium* in CD cases, species previously linked to pre-T1D (32), Behcet's disease (33), and obesity (34), respectively. A decrease in the abundance of a number of species and strains was observed, such as in *S. thermophilus*, *F. prausnitzii*, and *C. clostridioforme*, previously described as acting as a probiotic (36), blocking release of inflammatory cytokines (37, 38), and related to butyrate production (42). Microbes with increased abundance were detected that had not been reported before and are considered CD specific. A notable example is *Porphyromonas* sp., specifically strain 31_2, for which we observed an increase in abundance at all time points except -18 mo. Other species of *Porphyromonas*, such as *Porphyromonas gingivalis*, have previously been reported to lead to activation of the innate immune system (78) and impairment of the gut barrier (79, 80), in addition to contributing to development of rheumatoid arthritis and increased disease severity (78). The persistent increase in abundance of *Porphyromonas* sp. in cases prior to CD onset suggests that *Porphyromonas* sp. 31_2 is a strain that may contribute to CD pathogenesis. These previously undetected microbes highlight the importance of species and strain differences in the microbiome and metabolome of disease processes. In addition to these alterations in bacterial/archaeal species, we also identified changes in one protist species and one viral species at -15 mo. It should be noted that the low abundance of fungi, protists, and viruses compared with the bacterial/archaeal part of the gut microbiota (Dataset S2) may affect the detection of additional significant changes in them.

The analysis of pathways also revealed that the majority of altered pathways in cases were differentially abundant as early as -15 mo compared with CD onset, suggesting that a potential epigenetic influence of the gut microbiome on key functions, including immune functions dictating gluten tolerance/immune response balance, can start months before onset of the disease. The observation that the fluctuating presence of tTG antibodies can also occur months before onset of CD supports the hypothesis of a potential epigenetic mechanism.

We identified pathways for which there is conflicting evidence for a role in CD. We observed an increased abundance of tryptophan metabolism in cases at -15 mo compared with CD onset, but we also observed a lower abundance of tryptophan metabolism in cases at -18, -12, -6, and -3 mo compared with CD onset. Increased tryptophan metabolism has been linked to CD as increased

expression of enzymes catalyzing tryptophan degradation has been observed in duodenal tissue of patients with active CD (81). The increased activity of these enzymes has also been suggested to exert anti-inflammatory properties through decreasing the rate at which tissue is damaged following development of autoimmune disease (82, 83). We also observed decreased abundance of LPS biosynthesis pathways at all of the time points in cases except at ~ 15 mo compared with CD onset. Existing literature pertaining to LPS shows that microbially derived LPS may either stimulate or inhibit an immune response and consequently, lead to decreased or increased risk of developing autoimmune diseases (84–86). For example, LPS produced by *E. coli* has been shown to inhibit an immune response leading to decreased risk of T1D in mice (84), while LPS produced by *Bacteroides dorei* has been shown to contribute to T1D pathogenesis (84, 87). Our metabolite analysis identified serine, threonine, and 3-hydroxyphenylacetic acid, which were increased at all time points compared with CD onset, and they have been linked to the regulation of the adaptive immune response (51), rheumatoid arthritis (52), and decreased proinflammatory cytokine production (53).

Longitudinal analysis of microbial species, pathways, and metabolites in control subjects revealed alterations in the microbial species and metabolites linked to protection against chronic immune-based disorders. For example, control subjects were found to have an increased abundance of *E. eligens*, six strains of *B. longum*, and nine strains of *B. breve*. An increase in abundance of *B. longum* has been reported to be anti-inflammatory in animal models of gliadin-induced enteropathy (54, 55). Furthermore, previous work found a higher abundance of *B. longum* in control subjects in the first 6 mo after birth compared with subjects who later developed CD (17). *B. breve* has previously been shown to reduce production of serum proinflammatory cytokines when administered to patients newly diagnosed with CD (57), possibly induced by production of short-chain fatty acids (58). Pathway analysis revealed only minimal changes in pathways in controls, including a decreased abundance in sulfur metabolism, for which an increase has been associated with T1D (62). Metabolomic analysis also identified several metabolites of significance not previously associated with protection against autoimmune conditions.

Finally, we were able to link microbes, pathways, and metabolites identified as significant in cross-sectional and longitudinal analyses by performing association analyses (based on Spearman correlation) and unraveled microbiota–host metabolites interactions by using pathway enrichment analysis. These analyses provided insights into how interactions between microbes, pathways, and metabolites in the gut may modulate host pathways involved in inflammation and autoimmunity.

A potential limitation of this study is the low number of cases and controls, which limits the power of our statistical analysis. However, this study was intended to serve as a proof of concept pilot study, yet it allowed determination of significant microbiome alterations consonant with previously reported findings that linked specific microbes and metabolites to pro- or anti-inflammatory effects. This highlights the power of a longitudinal approach and underscores the need for further recruitment and sampling to increase sample size. The CDGEMM cohort is designed to address this limitation as we expect ~ 50 subjects to develop CD by the end of the study period (26), which will be used to validate our current findings and to refine our predictions that link specific microbiome/metabolome shifts to risk of developing CD. Another potential limitation of the study is the collection and analysis of fecal samples, which may not serve as an ideal measure of the small intestinal microbiota, the location of interest in CD. However, longitudinal sampling of the small intestine in cases and controls is not feasible nor justifiable in a birth cohort study. The comparison of the fecal and mucosal microbiome has been addressed by other groups (88–90) and is an approach that could potentially be performed in our cohort in subjects at CD onset.

Finally, it should be noted that while shotgun metagenomics is considered the state-of-the-art technology for microbiome profiling, the compatibility of metagenomic studies is currently hampered by inconsistencies and biases introduced by experimental protocols, data trimming approaches, and statistical parameters that are not sound descriptors of the underlying microbial community properties (91, 92). In this study, we minimized these biases by following the latest standards; however, using averages over multiple days of sampling instead of the single-day sampling that was conducted in this study could increase the reliability of our multiomics analyses results as has been highlighted previously (93). Additionally, since a large proportion of the short reads cannot be mapped to any microbial taxon or pathway in metagenomic analysis, our results do not capture the full complexity of the gut microbiota. Additionally, future work will include further analysis of these changes related to other influencing factors, such as timing of gluten introduction, amount of gluten consumed, and exposure to viruses. Identifying changes in the microbiome and metabolome as markers of progression toward CD onset will not only provide insight into the earliest steps in the pathogenesis of CD onset but also, provide tangible ways to manipulate the intestinal microbiota to prevent the disease. Since CD serves as a model of autoimmune diseases, the approach described here could pave the way for designing personalized prevention and treatment strategies for CD and potentially in the future, for other autoimmune conditions.

Conclusions

In this study, we presented a multiomics analysis of the gut microbiota in CD by using a prospective longitudinal birth cohort study design. Despite the limited sample size, our longitudinal analysis uncovered several species, pathways, and metabolites, for which the abundance is significantly altered before CD onset compared to disease onset and which is distinct from results of matched controls. The identified alterations point to a march from the preclinical stage of disease to a break of tolerance to gluten and subsequent onset of CD and may serve as microbial markers of progression toward disease onset. This onset is characterized by complex patterns of increased abundances of proinflammatory species and decreased abundances of protective and anti-inflammatory species at various time points preceding the onset of the disease. These microbiome shifts, coupled with metabolome findings, may represent potential biomarkers of CD development, which can be further scrutinized to pinpoint tractable points of intervention in the gut microbiota/metabolome to restore tolerance to gluten and prevent autoimmunity. Our study highlights the need for a transition from case-control to prospective longitudinal microbiome studies in order to uncover causal links between gut microbiota dysbiosis and disease pathogenesis that can usefully serve as predictive biomarkers of disease onset and possibly provide therapeutic or preventive targets to intercept CD before its onset.

Methods

Study Population. Ten infants who developed CD and their matched controls from the CDGEMM prospective birth cohort study (26) were selected for study. CDGEMM enrolls healthy infants between the ages of 0 to 6 mo who have a first-degree relative with CD and follows them prospectively for 5 y, with an additional 5 y until age 10 if parents elect. As part of the study, questionnaires are given to parents at enrollment to obtain information related to the infants' environment at birth and delivery; monthly diaries to monitor food intake and antibiotic exposure are also utilized. In addition to clinical information, stool samples were collected into four tubes every 3 mo from birth to 3 y of age and then every 6 mo until age 10 y. Two tubes of stool are stored in the United States and two are stored in Italy to allow for processing for both metagenomic and metabolomic analyses. Serum is collected every 6 mo until 3 y of age and then yearly until 5 y of age. All infants undergo serum testing for antibodies to immunoglobulin A (IgA) tTG and immunoglobulin G (IgG) deamidated gliadin peptide (dGP) using QUANTA Lite Rh-tTG IgA ELISA (INOVA Diagnostics) on the BioFlash platform at each collection. HLA genetic type is determined from whole blood collected at 12 mo of age using the DQ-CD Typing Plus (BioDiagne) per the manufacturer's instructions.

Infants found to have IgA tTG levels above the kit reference value (>20 chemiluminescent units) were subjected to confirmatory testing for IgA EMAs using the NOVA Lite Monkey Esophagus IFA Kit (Inova Diagnostics). Infants found to have elevated IgG dGP in the absence of elevated IgA tTG were evaluated for potential IgA deficiency. A total IgA-level measurement was performed for serum samples from these individuals using immuno-turbidimetric methods (LabCorp). Parents were informed of serology results after each blood draw and if positive, instructed to follow up with their physician for further confirmatory testing, including repeating blood work and upper endoscopy. Written informed consent was obtained from the parents of infants included in the study. This study was approved by the Partners Human Research Committee Institutional Review Board.

DNA Extraction. Stool samples for metagenomic analysis were stored and processed in the United States. Total DNA from each sample was extracted using the Qiagen Power soil DNA extraction kit (Qiagen).

Metagenomic Sequencing. Isolated DNA was quantified by Qubit 2.0 (ThermoFisher). DNA libraries were prepared using the Illumina Nextera XT library preparation kit according to the manufacturer's protocol. Library quantity and quality were assessed with Qubit (ThermoFisher) and Tapestation (Agilent Technologies). Libraries were then sequenced on the Illumina HiSeq 400 platform on a 2 × 150-bp run at CosmosID Inc.

Taxonomic Profiling. In accordance with our previously published work (19), sequence quality assessment and trimming of metagenomic reads were performed by using the MultiQC approach (79). Taxonomic profiling of metagenomics samples was performed at both species- and strain-level resolutions for bacteria, archaea, viruses, fungi, and protists using the commercial CosmosID (CosmosID Inc.) metagenomic analysis platform [formerly knowns as GENIUS (94, 95); <https://app.cosmosid.com/login>], which is based on an assembly-free kmer-based method. *SI Appendix, SI Text* has a description of this platform, and **Dataset S2** has information on the sequencing depth of each sample and the number of reads with a taxon assignment. Of note, the microbial strains identified using this framework and those reported in this manuscript may not represent the actual microbial strains that inhabit the intestines of the study subjects; rather, they represent microbial strains in the reference dataset that are closely related to those in the gut microbiota of these subjects.

Functional Profiling. Functional profiling to identify functional Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways encoded by each metagenome and their abundances was performed as described previously (19). Raw metagenomic reads were first trimmed using BBduk (<https://jgi.doe.gov/data-and-tools/bbtools/>; with parameters minlen = 25, qtrim = rl, trimq = 20). Next, we used SPADes (96) (with parameter *–only-assembler -k 77,99,127*) to perform metagenome assembly for each sample. After removing short contigs (length threshold = 500 base pairs), we used Prodigal (97) (v2.6 using *-d* parameter) to identify coding sequences in the assembled metagenomes and then Interproscan (97) (with parameters *-appl Hamap, ProDom -p, -f tsv*) to assign biochemical functions to identified genes/proteins based on KEGG pathways. The relative abundance of each gene was computed using $G = \frac{L \times C}{(R - K + 1)}$ where G is fragments per kilobase of gene per million, L is the length of the gene, C is the coverage of the contig on which the gene is identified, R is the read length, and K is the kmer size (98). The relative abundances of each KEGG pathway were then quantified by summing the relative abundances of all the genes associated with that pathway. **Dataset S4** shows the percentage of genes in each sample mapped to KEGG pathways.

Metabolomic Profiling. Metabolomic profiling of stool samples was performed in Italy as described in ref. 19 and detailed in *SI Appendix, SI Text*. Briefly, metabolome extraction, purification, and derivatization were performed by

using the MetaboPrep GC kit (Theoreo), and instrumental analyses were performed with a Gas chromatography–mass spectrometry (GC-MS) system (GC-2010 Plus gas chromatograph and QP2010 Plus mass spectrometer; Shimadzu Corp.). Sample analysis was also conducted in triplicate. The molecular identity of metabolites was determined by analyzing the corresponding mass spectrum in the chromatogram with the linear index difference max tolerance set to 10. These identified metabolites were further confirmed using external standards according to level 1 Metabolomics Standards Initiative (19, 99).

Cross-Sectional and Longitudinal Analyses. Cross-sectional analysis to identify species, strains, pathways, or metabolites whose abundance is significantly different between cases and controls at a given time point (i.e., CD onset) was performed as previously described (19) by using the Mann–Whitney *U* (Wilcoxon rank-sum) test (*P* value threshold of 0.05 was used to report significant results). Longitudinal differential abundance analysis of species, strains, pathways, or metabolites between each time point before CD onset (−18, −15, 12, −9, −6, and −3 mo) and the time of CD onset (*t* = 0) was performed by using paired Wilcoxon (Wilcoxon signed rank) test points with the same *P* value thresholds noted above to report the significant results. Any longitudinal pattern observed for both cases and controls was not reported. Analyses of microbial species, strains, and pathways were performed in Python (using *scipy.stats.mannwhitneyu* and *scipy.stats.wilcoxon* functions), and those for metabolites were performed in R [using the *ttest.Anal* function of the MetaboAnalyst 4.0 (100) using parameters *nonpar = TRUE* and *paired = FALSE* for the cross-sectional analysis and *paired = TRUE* for the longitudinal analysis].

Pathway Enrichment Analysis. Metabolites identified as significant either in cross-sectional or in longitudinal analyses were subjected to pathway enrichment analysis (based on human pathways) with MetaboAnalyst 4.0 (100), using the Metabolite Set Enrichment Analysis module. Accession numbers of detected metabolites (i.e., human metabolome database identification) were generated, manually inspected, and utilized to map the canonical pathways.

Declarations: Ethics Approval and Consent to Participate. Written informed consent was obtained from the parents of infants included in the study according to the standards outlined and approved by the Partners Human Research Committee Institutional Review Board.

Data Availability. Raw sequence data generated in this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (SRA) repository (BioProjectID [PRJNA486782](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA486782) and SRA accession no. [SRP158417](https://www.ncbi.nlm.nih.gov/sra/SRP158417)). All other data from the analyses are included in the manuscript and/or [supporting information](#).

ACKNOWLEDGMENTS. We thank the families that participate in this study and whose contribution was instrumental to the findings described in this manuscript and the CDGEMM team, including Monica Montuori, Pasqua Piemontese, Angela Calvi, Mariella Baldassarre, Lorenzo Norsa, Celeste Lidia Raguseo, Tiziana Passaro, Paola Roggero, Marco Crocco, Annalisa Morelli, Michela Perrone, Naire Sansotta, Marcello Chieppa, Giovanni Scala, Maria Elena Lionetti, Carlo Catassi, Adelaide Serretiello, Corrado Vecchi, and Gemma Castillejo de Villante. This work was partially supported by funding from NIH, National Institute of Diabetes and Digestive and Kidney Diseases Grants DK109620 (to M.M.L.), K23DK122127 (to M.M.L.), and DK104344 (to A.F.); funding from Nutrition Obesity Research Center at Harvard Grant P30-DK040561 (to M.M.L.); the Thrasher Research Fund (M.M.L.); the faculty start-up funding by the Mucosal Immunology and Biology Research Center at Massachusetts General Hospital (A.R.Z.); and the support of Joyce and Hugh McCormick and Hilary and Langley Steinert. Partial funding for R.C. is from NSF Grant CCF1918749.

1. E. Lionetti, S. Gatti, A. Pulvirenti, C. Catassi, Celiac disease from a global perspective. *Best Pract. Res. Clin. Gastroenterol.* **29**, 365–379 (2015).
2. I. Ricaño-Ponce, C. Wijmenga, J. Gutierrez-Achury, Genetics of celiac disease. *Best Pract. Res. Clin. Gastroenterol.* **29**, 399–412 (2015).
3. B. Jabri, D. D. Kasarda, P. H. Green, Innate and adaptive immunity: The yin and yang of celiac disease. *Immunol. Rev.* **206**, 219–231 (2005).
4. S. M. Kim, T. Mayassi, B. Jabri, Innate immunity: Actuating the gears of celiac disease pathogenesis. *Best Pract. Res. Clin. Gastroenterol.* **29**, 425–435 (2015).
5. M. M. Leonard *et al.*, RNA sequencing of intestinal mucosa reveals novel pathways functionally linked to celiac disease pathogenesis. *PLoS One* **14**, e0215132 (2019).
6. G. Serena *et al.*, Proinflammatory cytokine interferon- γ and microbiome-derived metabolites dictate epigenetic switch between forkhead box protein 3 isoforms in coeliac disease. *Clin. Exp. Immunol.* **187**, 490–506 (2017).
7. G. Serena *et al.*, Intestinal epithelium modulates macrophage response to gliadin in celiac disease. *Front. Nutr.* **6**, 167 (2019).
8. M. Nanayakkara *et al.*, P31-43, an undigested gliadin peptide, mimics and enhances the innate immune response to viruses and interferes with endocytic trafficking: A role in celiac disease. *Sci. Rep.* **8**, 10821 (2018).
9. S. Tamburini, N. Shen, H. C. Wu, J. C. Clemente, The microbiome in early life: Implications for health outcomes. *Nat. Med.* **22**, 713–722 (2016).
10. X. C. Morgan *et al.*, Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* **13**, R79 (2012).
11. C. J. Stewart *et al.*, Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).
12. S. Jangi *et al.*, Alterations of the human gut microbiome in multiple sclerosis. *Nat. Commun.* **7**, 12015 (2016).

13. G. De Palma *et al.*, Interplay between human leukocyte antigen genes and the microbial colonization process of the newborn intestine. *Curr. Issues Mol. Biol.* **12**, 1–10 (2010).

14. M. Olivares *et al.*, The HLA-DQ2 genotype selects for early intestinal microbiota composition in infants at high risk of developing coeliac disease. *Gut* **64**, 406–417 (2015).

15. M. Sellitto *et al.*, Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants. *PLoS One* **7**, e33837 (2012).

16. T. Pozo-Rubio *et al.*, Influence of early environmental factors on lymphocyte subsets and gut microbiota in infants at risk of celiac disease; The PROFICEL study. *Nutr. Hosp.* **28**, 464–473 (2013).

17. M. Olivares *et al.*, Gut microbiota trajectory in early life may predict development of celiac disease. *Microbiome* **6**, 36 (2018).

18. A. Rintala *et al.*, Early fecal microbiota composition in children who later develop celiac disease and associated autoimmunity. *Scand. J. Gastroenterol.* **53**, 403–409 (2018).

19. M. M. Leonard *et al.*; CD-GEMM Team, Multi-omics analysis reveals the influence of genetic and environmental risk factors on developing gut microbiota in infants at risk of celiac disease. *Microbiome* **8**, 130 (2020).

20. Q. Huang *et al.*, Children developing celiac disease have a distinct and proinflammatory gut microbiota in the first 5 years of life. <https://doi.org/10.1101/2020.02.29.971242> (5 March 2020).

21. K. Zafeiropoulou *et al.*, Alterations in intestinal microbiota of children with celiac disease at the time of diagnosis and on a gluten-free diet. *Gastroenterology* **159**, 2039–2051.e20 (2020).

22. A. R. Di Biase *et al.*, Gut microbiota signatures and clinical manifestations in celiac disease children at onset: A pilot study. *J. Gastroenterol. Hepatol.* **36**, 446–454 (2021).

23. M. M. Leonard, A. Fasano, The microbiome as a possible target to prevent celiac disease. *Expert Rev. Gastroenterol. Hepatol.* **10**, 555–556 (2016).

24. B. B. Warner *et al.*, Gut bacteria dysbiosis and necrotising enterocolitis in very low birthweight infants: A prospective case-control study. *Lancet* **387**, 1928–1936 (2016).

25. W. Turpin *et al.*; Crohn's and Colitis Canada Genetic Environmental Microbial Project Research Consortium; CCC GEM Project recruitment site directors include Maria Abreu, Kenneth Croitoru, Increased intestinal permeability is associated with later development of Crohn's disease. *Gastroenterology* **159**, 2092–2100.e5 (2020).

26. M. M. Leonard, S. Camhi, T. B. Huedo-Medina, A. Fasano, Celiac disease genomic, environmental, microbiome, and metabolomic (CDGEMM) study design: Approach to the future of personalized prevention of celiac disease. *Nutrients* **7**, 9325–9336 (2015).

27. S. Husby *et al.*, European Society Paediatric Gastroenterology, Hepatology and Nutrition guidelines for diagnosing coeliac disease 2020. *J. Pediatr. Gastroenterol. Nutr.* **70**, 141–156 (2020).

28. N. Yoshida *et al.*, *Bacteroides vulgatus* and *Bacteroides dorei* reduce gut microbial lipopolysaccharide production and inhibit atherosclerosis. *Circulation* **138**, 2486–2498 (2018).

29. O. Cinek *et al.*, Imbalance of bacteriome profiles within the Finnish diabetes prediction and prevention study: Parallel use of 16S profiling and virome sequencing in stool samples from children with islet autoimmunity and matched controls. *Pediatr. Diabetes* **18**, 588–598 (2017).

30. P. Gauffin Cano, A. Santacruz, Á. Moya, Y. Sanz, *Bacteroides uniformis* CECT 7771 ameliorates metabolic and immunological dysfunction in mice with high-fat-diet induced obesity. *PLoS One* **7**, e41079 (2012).

31. A. Haghikia *et al.*, Dietary fatty acids directly impact central nervous system autoimmunity via the small intestine. *Immunity* **43**, 817–829 (2015).

32. C. Maffei *et al.*, Association between intestinal permeability and faecal microbiota composition in Italian children with beta cell autoimmunity at risk for type 1 diabetes. *Diabetes Metab. Res. Rev.* **32**, 700–709 (2016).

33. Z. Ye *et al.*, A metagenomic study of the gut microbiome in Behcet's disease. *Microbiome* **6**, 135 (2018).

34. K. Kameyama, K. Itoh, Intestinal colonization by a *Lachnospiraceae* bacterium contributes to the development of diabetes in obese mice. *Microbes Environ.* **29**, 427–430 (2014).

35. Y. Nakanishi, T. Sato, T. Ohteki, Commensal Gram-positive bacteria initiates colitis by inducing monocyte/macrophage mobilization. *Mucosal Immunol.* **8**, 152–160 (2015).

36. S. Ménard *et al.*, Lactic acid bacteria secrete metabolites retaining anti-inflammatory properties after intestinal transport. *Gut* **53**, 821–828 (2004).

37. H. Sokol *et al.*, *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16731–16736 (2008).

38. E. Quévrain *et al.*, Identification of an anti-inflammatory protein from *Faecalibacterium prausnitzii*, a commensal bacterium deficient in Crohn's disease. *Gut* **65**, 415–425 (2016).

39. H. Sokol *et al.*, Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm. Bowel Dis.* **15**, 1183–1189 (2009).

40. K. Machiels *et al.*, A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* **63**, 1275–1283 (2014).

41. R. Pittayanan *et al.*, Differences in gut microbiota in patients with vs without inflammatory bowel diseases: A systematic review. *Gastroenterology* **158**, 930–946.e1 (2020).

42. P. Dehoux *et al.*, Comparative genomics of *Clostridium bolteae* and *Clostridium clostridioforme* reveals species-specific genomic properties and numerous putative antibiotic resistance determinants. *BMC Genomics* **17**, 819 (2016).

43. A. Fite *et al.*, Longitudinal analyses of gut mucosal microbiotas in ulcerative colitis in relation to patient age and disease severity and duration. *J. Clin. Microbiol.* **51**, 849–856 (2013).

44. M.-W. Chien, S. H. Fu, C. Y. Hsu, Y. W. Liu, H. K. Sytwu, The modulatory roles of N-glycans in T-cell-mediated autoimmune diseases. *Int. J. Mol. Sci.* **19**, 780 (2018).

45. Z. Yu *et al.*, Family studies of type 1 diabetes reveal additive and epistatic effects between MGAT1 and three other polymorphisms. *Genes Immun.* **15**, 218–223 (2014).

46. A. M. Dias *et al.*, Dysregulation of T cell receptor N-glycosylation: A molecular mechanism involved in ulcerative colitis. *Hum. Mol. Genet.* **23**, 2416–2427 (2014).

47. H. Mkhitaryan *et al.*, Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. *Nat. Commun.* **2**, 334 (2011).

48. R. Morgan *et al.*, N-acetylglucosaminyltransferase V (Mgat5)-mediated N-glycosylation negatively regulates Th1 cytokine production by T cells. *J. Immunol.* **173**, 7200–7208 (2004).

49. K. S. Lau *et al.*, Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. *Cell* **129**, 123–134 (2007).

50. M. Orešić *et al.*, Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. *J. Exp. Med.* **205**, 2975–2984 (2008).

51. E. H. Ma *et al.*, Serine is an essential metabolite for effector T cell expansion. *Cell Metab.* **25**, 345–357 (2017).

52. R. K. Madsen *et al.*, Diagnostic properties of metabolic perturbations in rheumatoid arthritis. *Arthritis Res. Ther.* **13**, R19 (2011).

53. M. Monagas *et al.*, Dihydroxylated phenolic acids derived from microbial metabolism reduce lipopolysaccharide-stimulated cytokine secretion by human peripheral blood mononuclear cells. *Br. J. Nutr.* **102**, 201–206 (2009).

54. J. M. Laparra, M. Olivares, O. Gallina, Y. Sanz, *Bifidobacterium longum* CECT 7347 modulates immune responses in a gliadin-induced enteropathy animal model. *PLoS One* **7**, e30744 (2012).

55. M. Medina, G. De Palma, C. Ribes-Koninckx, M. Calabuig, Y. Sanz, *Bifidobacterium* strains suppress *in vitro* the pro-inflammatory milieu triggered by the large intestinal microbiota of coeliac patients. *J. Inflamm. (Lond.)* **5**, 19 (2008).

56. C. B. Wong, N. Iwabuchi, J. Z. Xiao, Exploring the science behind *Bifidobacterium breve* M-16V in infant health. *Nutrients* **11**, 1724 (2019).

57. M. Klemenak, J. Dolinšek, T. Langerholc, D. Di Gioia, D. Mičetić-Turk, Administration of *Bifidobacterium breve* decreases the production of TNF- α in children with celiac disease. *Dig. Dis. Sci.* **60**, 3386–3392 (2015).

58. M. Primec *et al.*, Clinical intervention using *Bifidobacterium* strains in celiac disease children reveals novel microbial modulators of TNF- α and short-chain fatty acids. *Clin. Nutr.* **38**, 1373–1381 (2019).

59. J. K. Maerz *et al.*, Bacterial immunogenicity is critical for the induction of regulatory B cells in suppressing inflammatory immune responses. *Front. Immunol.* **10**, 3093 (2020).

60. W. S. F. Chung *et al.*, Prebiotic potential of pectin and pectic oligosaccharides to promote anti-inflammatory commensal bacteria in the human colon. *FEMS Microbiol. Ecol.* **93**, fix127 (2017).

61. J. Shimizu *et al.*, Propionate-producing bacteria in the intestine may associate with skewed responses of IL10-producing regulatory T cells in patients with relapsing polychondritis. *PLoS One* **13**, e0203657 (2018).

62. C. T. Brown *et al.*, Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. *PLoS One* **6**, e25792 (2011).

63. Y. Wei *et al.*, Alterations of gut microbiome in autoimmune hepatitis. *Gut* **69**, 569–577 (2020).

64. E. Sánchez, E. Donat, C. Ribes-Koninckx, M. Calabuig, Y. Sanz, Intestinal *Bacteroides* species associated with coeliac disease. *J. Clin. Pathol.* **63**, 1105–1111 (2010).

65. Y. Murakami, S. Hanazawa, K. Nishida, H. Iwasaka, S. Kitano, N-acetyl-D-galactosamine inhibits TNF- α gene expression induced in mouse peritoneal macrophages by fimbriae of Porphyromonas (*Bacteroides*) gingivalis, an oral anaerobe. *Biochem. Biophys. Res. Commun.* **192**, 826–832 (1993).

66. M. J. Acosta *et al.*, Coenzyme Q biosynthesis in health and disease. *Biochim. Biophys. Acta* **1857**, 1079–1085 (2016).

67. A. Mohanty, R. Tiwari-Pandey, N. R. Pandey, Mitochondria: The indispensable players in innate immunity and guardians of the inflammatory response. *J. Cell Commun. Signal.* **13**, 303–318 (2019).

68. Y. Oishi *et al.*, SREBP1 contributes to resolution of pro-inflammatory TLR4 signaling by reprogramming fatty acid metabolism. *Cell Metab.* **25**, 412–427 (2017).

69. A. Perl, Review: Metabolic control of immune system activation in rheumatic diseases. *Arthritis Rheumatol.* **69**, 2259–2270 (2017).

70. N. Wegner, R. Wait, P. J. Venables, Evolutionarily conserved antigens in autoimmune disease: Implications for an infective aetiology. *Int. J. Biochem. Cell Biol.* **41**, 390–397 (2009).

71. J. D. Watkins *et al.*, "Aminoacyl tRNA Synthetases for Modulating Inflammation." Google Patent US9127268B2 (2018).

72. T. Vatanen *et al.*, The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature* **562**, 589–594 (2018).

73. L. A. DiMeglio, C. Evans-Molina, R. A. Oram, Type 1 diabetes. *Lancet* **391**, 2449–2462 (2018).

74. E. Lionetti *et al.*; SIGENP (Italian Society of Pediatric Gastroenterology, Hepatology, and Nutrition) Working Group on Weaning and CD Risk, Introduction of gluten, HLA status, and the risk of celiac disease in children. *N. Engl. J. Med.* **371**, 1295–1303 (2014).

75. S. L. Vriezinga *et al.*, Randomized feeding intervention in infants at high risk for celiac disease. *N. Engl. J. Med.* **371**, 1304–1315 (2014).

76. K. Karlinger, T. Györke, E. Makó, A. Mester, Z. Tarján, The epidemiology and the pathogenesis of inflammatory bowel disease. *Eur. J. Radiol.* **35**, 154–167 (2000).

77. A. J. Thompson, S. E. Baranzini, J. Geurts, B. Hemmer, O. Ciccarelli, Multiple sclerosis. *Lancet* **391**, 1622–1636 (2018).

78. M. Pouliot, C. B. Clish, N. A. Petasis, T. E. Van Dyke, C. N. Serhan, Lipoxin A(4) analogues inhibit leukocyte recruitment to *Porphyromonas gingivalis*: A role for cyclooxygenase-2 and lipoxins in periodontal disease. *Biochemistry* **39**, 4761–4768 (2000).

79. K. Sato *et al.*, Aggravation of collagen-induced arthritis by orally administered *Porphyromonas gingivalis* through modulation of the gut microbiota and gut immune system. *Sci. Rep.* **7**, 6955 (2017).

80. M. B. Flak *et al.*, Inflammatory arthritis disrupts gut resolution mechanisms, promoting barrier breakdown by *Porphyromonas gingivalis*. *JCI Insight* **4**, e125191 (2019).

81. M. I. Torres, M. A. López-Casado, P. Lorite, A. Ríos, Tryptophan metabolism and indoleamine 2,3-dioxygenase expression in celiac disease. *Clin. Exp. Immunol.* **148**, 419–424 (2007).

82. A. L. Mellor, D. H. Munn, Tryptophan catabolism and T-cell tolerance: Immunosuppression by starvation? *Immunol. Today* **20**, 469–473 (1999).

83. J. Gao *et al.*, Impact of the gut microbiota on intestinal immunity mediated by tryptophan metabolism. *Front. Cell. Infect. Microbiol.* **8**, 13 (2018).

84. T. Vatanen *et al.*; DIABIMMUNE Study Group, Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* **165**, 842–853 (2016).

85. N. A. Granholm, T. Cavallo, Long-lasting effects of bacterial lipopolysaccharide promote progression of lupus nephritis in NZB/W mice. *Lupus* **3**, 507–514 (1994).

86. M. Murakami *et al.*, Oral administration of lipopolysaccharides activates B-1 cells in the peritoneal cavity and lamina propria of the gut and induces autoimmune symptoms in an autoantibody transgenic mouse. *J. Exp. Med.* **180**, 111–121 (1994).

87. A. G. Davis-Richardson *et al.*, *Bacteroides dorei* dominates gut microbiome prior to autoimmunity in Finnish children at high risk for type 1 diabetes. *Front. Microbiol.* **5**, 678 (2014).

88. S. Panelli *et al.*, Comparative study of salivary, duodenal, and fecal microbiota composition across adult celiac disease. *J. Clin. Med.* **9**, 1109 (2020).

89. R. Di Cagno *et al.*, Duodenal and faecal microbiota of celiac children: Molecular, phenotype and metabolome characterization. *BMC Microbiol.* **11**, 219 (2011).

90. M. C. Collado, E. Donat, C. Ribes-Koninckx, M. Calabuig, Y. Sanz, Specific duodenal and faecal bacterial groups associated with paediatric coeliac disease. *J. Clin. Pathol.* **62**, 264–269 (2009).

91. S. Nayfach, K. S. Pollard, Toward accurate and quantitative comparative metagenomics. *Cell* **166**, 1103–1116 (2016).

92. C. Poussin *et al.*, Interrogating the microbiome: Experimental and computational considerations in support of study reproducibility. *Drug Discov. Today* **23**, 1644–1657 (2018).

93. M. Poyet *et al.*, A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat. Med.* **25**, 1442–1452 (2019).

94. N. A. Hasan *et al.*, Microbial community profiling of human saliva using shotgun metagenomic sequencing. *PLoS One* **9**, e97699 (2014).

95. D. Ponnusamy *et al.*, Cross-talk among flesh-eating *Aeromonas hydrophila* strains in mixed infection leading to necrotizing fasciitis. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 722–727 (2016).

96. A. Bankevich *et al.*, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

97. D. Hyatt *et al.*, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* **11**, 119 (2010).

98. D. R. Zerbino, E. Birney, Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2004).

99. L. W. Sumner *et al.*, Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics* **3**, 211–221 (2007).

100. J. Chong *et al.*, MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46** (W1), W486–W494 (2018).