nature communications

Article

https://doi.org/10.1038/s41467-024-52561-6

Gut microbiota wellbeing index predicts overall health in a cohort of 1000 infants

Received: 4 March 2024

Accepted: 10 September 2024

Published online: 27 September 2024

Check for updates

Brandon Hickman[1] ✉, Anne Salonen[1], Alise J. Ponsero[1], Roosa Jokela[1], Kaija-Leena Kolho[1,2], Willem M. de Vos[1,3] & Katri Korpela[1] ✉

The human gut microbiota is central in regulating all facets of host physiology, and in early life it is thought to influence the host's immune system and metabolism, affecting long-term health. However, longitudinally monitored cohorts with parallel analysis of faecal samples and health data are scarce. In our observational study we describe the gut microbiota development in the first 2 years of life and create a gut microbiota wellbeing index based on the microbiota development and health data in a cohort of nearly 1000 infants using clustering and trajectory modelling. We show that infants' gut microbiota development is highly predictable, following one of five trajectories, dependent on infant exposures, and predictive of later health outcomes. We characterise the natural healthy gut microbiota trajectory and several different dysbiotic trajectories associated with different health outcomes. *Bifidobacterium* and *Bacteroides* appear as early keystone organisms, directing microbiota development and consistently predicting positive health outcomes. A microbiota wellbeing index, based on the healthy development trajectory, is predictive of general health over the first 5 years. The results indicate that gut microbiota succession is part of infant physiological development, predictable, and malleable. This information can be utilised to improve the predictions of individual health risks.

The human gut microbiota plays an essential role in health, and in early life the gut microbiota is thought to influence the programming of the host's immune system, metabolism, and even nervous system[1–4]. Initial colonization of the human gut by bacteria occurs at birth by maternal gut microbes[5–7] and further develops through ecological succession in the first years of life[8,9]. It's been shown that successional stages in gut microbiota appear at specific time windows in infant development in a strikingly similar manner throughout the world, suggesting that humans may have adapted to relying on these age-dependent microbial signals as a guide to physiological maturation[8,10]. Indeed, several studies have shown that early life gut microbiota composition is predictive of later immune health[2,11–13] and adiposity[14–16], and studies in animal models have shown these effects to be causal[17–19].

Early life gut microbiota composition and development is influenced by many factors, notably birth mode, diet, and antibiotic use[8,20–23]. Many infants are subjected to microbiota-disrupting treatments or develop chronic diseases in early life and hence the gut microbiota development in such infants may not represent the optimal healthy pattern[24]. Furthermore, many studies that addressed the longitudinal microbiota development suffered from limited temporal resolution or did not follow in parallel the child's health[2,20,25] Consequently, the natural, undisturbed trajectory of microbiota development has not been thoroughly described[20,25]. It is not known if a single healthy development pattern can be identified, or whether different developmental patterns can lead to the same health outcomes. In addition, the relative contributions of initial gut microbiota

[1]Human Microbiome Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland. [2]Children's Hospital, University of Helsinki, Stenbäckinkatu 11, FI-00029 HUS Helsinki, Finland. [3]Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands.
✉e-mail: brandon.hickman@helsinki.fi; katri.korpela@helsinki.fi

colonisation (i.e., priority effects), postnatal exposures, and stochasticity, which in gut microbiota studies is often assumed to be strong, on the individual microbiota development have not been established.

Specific microbial taxa have been linked with various health outcomes in infants[26,27]. However, it is commonly postulated that the impact of gut microbiota on host health is not dependent on an individual taxon but on the overall community structure[28]. While often used, the term dysbiosis has not been defined in infants due to lack of data to support the long-term health associations of specific community types. We hypothesized that gut microbiota succession is part of an infant's physiological development, linked with the maturation of the immune system, closely associated with nutrition, metabolism, and growth, and may thus serve as an indicator of overall health and wellbeing in infants. Altered microbiota maturity indices in infancy have been associated to adverse health outcomes[29,30], but without reference to a normal healthy development, they are difficult to interpret. By utilising the Helmi cohort[31], a large longitudinal birth cohort of 984 infants monitored for associated health outcomes until 5 years of age, and 6203 faecal samples collected within the first two years of life, we describe a trajectory of undisturbed natural gut microbiota development in infants and create an index of gut microbiota wellbeing as an indicator of healthy infant development, thus defining "eubiosis" and "dysbiosis" in a large cohort.

Results

Impact of exposures and priority effect on infant gut microbiota

Utilising the 16S rRNA gene amplicon sequences of 984 infants (50.7% male; 49.3% female), 768 mothers, and 515 fathers, comprising 7211 faecal samples, in unsupervised principal coordinates analysis using log-Pearson distance (Fig. 1a), we observed a clear age gradient in the gut microbiota composition, with the infant gut microbiota gradually approaching, but not reaching, the adult-like composition over the first two years of life (age $p = 0.001$, $R^2 = 0.18$; 104 weeks vs adult $p = 0.001$, $R^2 = 0.21$). The infant gut microbiota composition varied greatly between individuals in the first 6 months (26 weeks) but converged thereafter (Fig. 1a–c). Birth mode was associated with the development (Fig. 1b, c). In the vaginally born infants not exposed to intrapartum antibiotics (VD), we observed an increase in principal component (PC) 2 scores in the first 26 weeks, and an increase in PC1 scores from 26 to 104 weeks (Fig. 1b, c). C-section birth (CS), and to a lesser extent, intrapartum antibiotic exposure during vaginal birth (V-ABX), altered this pattern, as these infants had consistently high PC2 scores already at 3 weeks (CS compared to VD, $p < 0.001$ at 3–78 weeks; $p = 0.003$ at 104 weeks), but rather showed a constant increase in PC1 scores already during the early weeks (PC 1 scores CS compared to VD, $p < 0.01$).

To identify factors (Supplementary Data 1) influencing the gut microbiota, we analysed the associations of background variables with the ordination (envfit) and performed permutational multivariate analysis of variance (adonis2) (Fig. 1d). The major determinant of PC1 scores was age ($R^2 = 0.178$ in adonis2). In addition, solid food consumption ($R^2 = 0.02$), use of antibiotics ($R^2 = 0.0008$), time spent with non-parental carer ($R^2 = 0.005$), and use of gut-targeting medications (laxative/antiflatulence/antidiarrheal/constipation; $R^2 = 0.004$ each) were significantly associated with PC1, after adjusting for age. Breastfeeding ($R^2 = 0.004$) and probiotic consumption had the opposite association, but the latter was not significant in a multivariate model ($p = 0.99$). The major driver of PC2 scores was birth mode ($R^2 = 0.02$) and having siblings ($R^2 = 0.007$).

To assess the predictability of microbiota development, we built multivariate regression models for the PC scores by age and quantified the variance explained by each factor at each time point. The most important determinant of gut microbiota composition was birth mode in the first 26 weeks, defecation rate throughout the first year, and diet and family composition from 1 to 2 years (52–104 weeks) (Fig. 2a). Maternal characteristics had a modest and consistent impact at all time points, while maternal microbiota composition (mother's PC scores) became influential at 26 weeks (1.22% of variation explained, $p = 0.017$) increasing with time (5.12% at 104 weeks, $p < 0.0001$) (Fig. 2a).

We added the infants' previous time points' microbiota composition (PC scores) into the models, discovering that it was by far the most influential factor determining the current composition, explaining 59% of the variation in the 6 weeks' samples (Fig. 2b, Supplementary Fig. 1a, $p < 0.001$). With the previous composition included, the impact of birth mode decreased or disappeared, demonstrating that the effect of birth mode on gut microbiota is due to its effects on the initial colonization. The impact of the initial composition was tested by replacing the previous time points' PC scores with the 3 weeks' PC scores (Fig. 2c, Supplementary Fig. 1b). The effect of the initial composition was strong for the first 26 weeks (over 10%), and still significant (1.7%) at 52 weeks ($p < 0.001$), highlighting the long-term importance of initial inoculation (Fig. 2b, c, Supplementary Fig. 1a, b).

The predictability of gut microbiota development was tested using model-based simulation in infants with full time series and complete background information ($N = 98$). Based on the background information (Fig. 2b) and the microbiota composition at 3 weeks, we used parameter estimates from the model that included the previous time points' microbiota composition to simulate the PC scores of each infant over time. The simulated microbiota development followed strikingly close to the observed patterns, achieving a correlation of 0.84, showing that the early life microbiota development is highly predictable with moderate stochasticity.

The impact of exposures was further assessed at each time point by projecting them onto the age-specific ordination using envfit. High maternal BMI, long duration of ruptured membranes, high gestational age, and formula feeding in parallel to high defecation rate had a similar but weaker impact on the gut microbiota as CS and V-ABX in the first 26 weeks (Supplementary Fig. 2). Maternal and infant probiotic use prior to sample collection had an impact like that of breastfeeding, while high appetite, siblings, and pacifier use were associated with formula-fed-like microbiota composition at 9 months (36 weeks) and beyond (Supplementary Fig. 2). Some associations may have been driven partly by confounders. For example, defecation rate, indicative of transit time, was strongly associated with breastfeeding—breastfeeding correlated with high defecation rate at 6 ($p = 0.012$) and 12 ($p = 0.017$) weeks but had the opposite association at 36 ($p = 0.035$), 78 ($p = 0.007$), and 104 ($p = 0.006$) weeks. Infants with a high appetite were more likely fed formula ($p < 0.05$) at 78 weeks (18 months). The PC scores were correlated with the individual microbial taxa to identify indicator organisms of overall microbial composition. These were *Bifidobacterium* at 3–26 weeks, *Bacillus* at 39 weeks, *Collinsella* at 12 months (52 weeks), *Enterococcus* at 78 weeks, and *Christensenella* at 104 weeks (Supplementary Fig. 2).

We then looked further into the taxonomic associations of the most important exposures. The abundant microbial families could be divided into those that naturally decline with age (*Bifidobactericeae, Bacteroidaceae, Enterobactericeae)*, peak at 26–52 weeks (*Veillonellaceae)*, or increase with age (*Lachnospiraceae, Ruminococcaceae*) (Fig. 2d). These patterns were affected by the infant's exposures. CS birth was strongly associated ($p < 0.001$) with delayed colonization by the genera *Bacteroides, Parabacteroides, Bifidobacterium*, and *Collinsella* (Fig. 2d, Supplementary Data 2). V-ABX had a CS-like impact mainly on the Gram-positive organisms (Fig. 2d). Breastfeeding was associated with increased relative abundance of *Lactobacillaceae* at 3 and 26–52 weeks ($p < 0.001$), *Bacteroidaceae* at 6 weeks ($p < 0.001$), and *Bifidobacteriaceae* at 36–52 weeks ($p < 0.01$), and a consistent decrease in *Lachnospiraceae* in the first two years of life ($p < 0.001$). Exclusive formula feeding was associated with advanced microbiota maturation indicated by faster *Enterobacteriaceae* decline, earlier *Veillonellaceae* peak and earlier increase in *Lachnospiraceae* (Fig. 2d,

https://doi.org/10.1038/s41467-024-52561-6

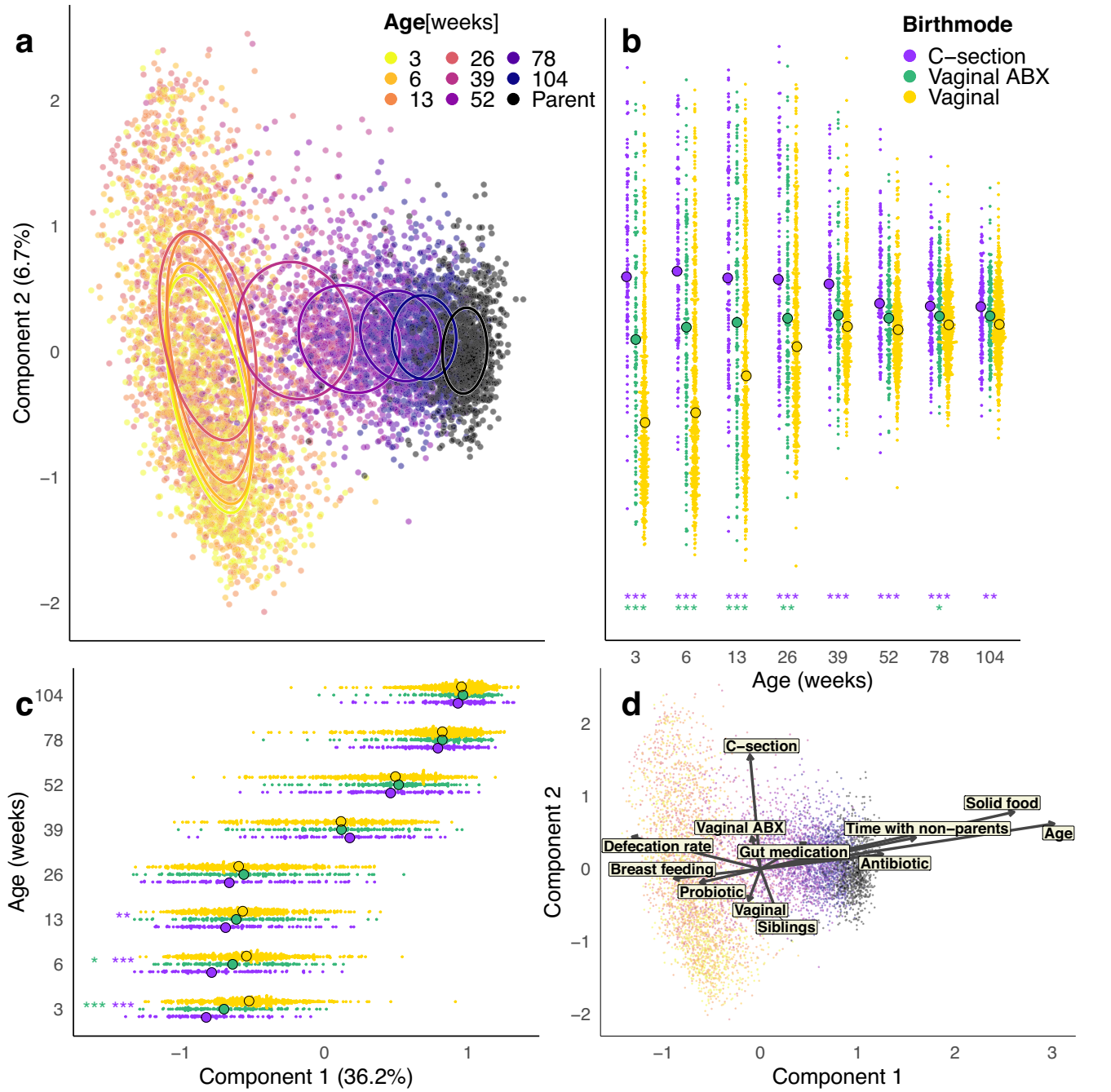

Fig. 1 | Infant gut microbiota in 984 infants ranging from 3 weeks to 2 years. **a** Principal coordinates (PC) analysis on Pearson correlation distances of log-transformed relative abundances of microbial genera. Colour shows the infant age from 3 to 104 weeks. Parent microbiota is represented as black. **b** PC Component 2 score against age and birth mode. Group median values are represented by the large circles. **c** PC Component 1 score against age and birth mode. Comparisons of birthmodes (C-section and Vaginal ABX to vaginally born infants) are represented with *P*-values are provided are colour-coded by comparison at the bottom (**b**) / left (**c**) of the panel based on a two-sided *T*-test. **d** Significant associations between the PC ordination and factors influencing gut microbiota according to envfit.

Supplementary Data 3). In the first weeks of life, having siblings was associated with increased relative abundance of *Bifidobacteriaceae* ($p < 0.0001$) and *Lactobacillaceae* ($p = 0.007$, Fig. 2d). The sibling effect on *Bifidobacteriaceae* was evident already at 3 weeks in the vaginally born ($p < 0.0001$), but not in the CS born ($p = 0.736$, Supplementary Data 4), suggesting that the effect is mediated by maternal gut microbiota. Indeed, we found that multiparous mothers had significantly higher relative faecal abundance of *Bifidobacterium* compared to primiparous mothers ($p < 0.05$). After 26 weeks, siblings had a more widespread impact, being associated with increased relative abundance of *Ruminococcaceae* and decreased *Lachnospiraceae* (Fig. 2d). In addition, green stool colour was associated with reduced relative abundance of *Bifidobacterium* at 6 and 13 weeks ($p = 0.007$).

The most important microbial taxa affecting gut microbiota development at each time point were identified using multivariate analysis of variance with the previous time points' taxonomic composition as the explanatory variables. *Bifidobacterium* and *Bacteroides* were the most influential taxa throughout the first 2 years, having an especially strong influence in the first months, while *Veillonella* and *Collinsella* became dominant influencers at 52–78 weeks (Supplementary Fig. 3). In the first 26 weeks, the relative abundances of *Bifidobacteriaceae* and *Bacteroidaceae* at a given time point were

https://doi.org/10.1038/s41467-024-52561-6

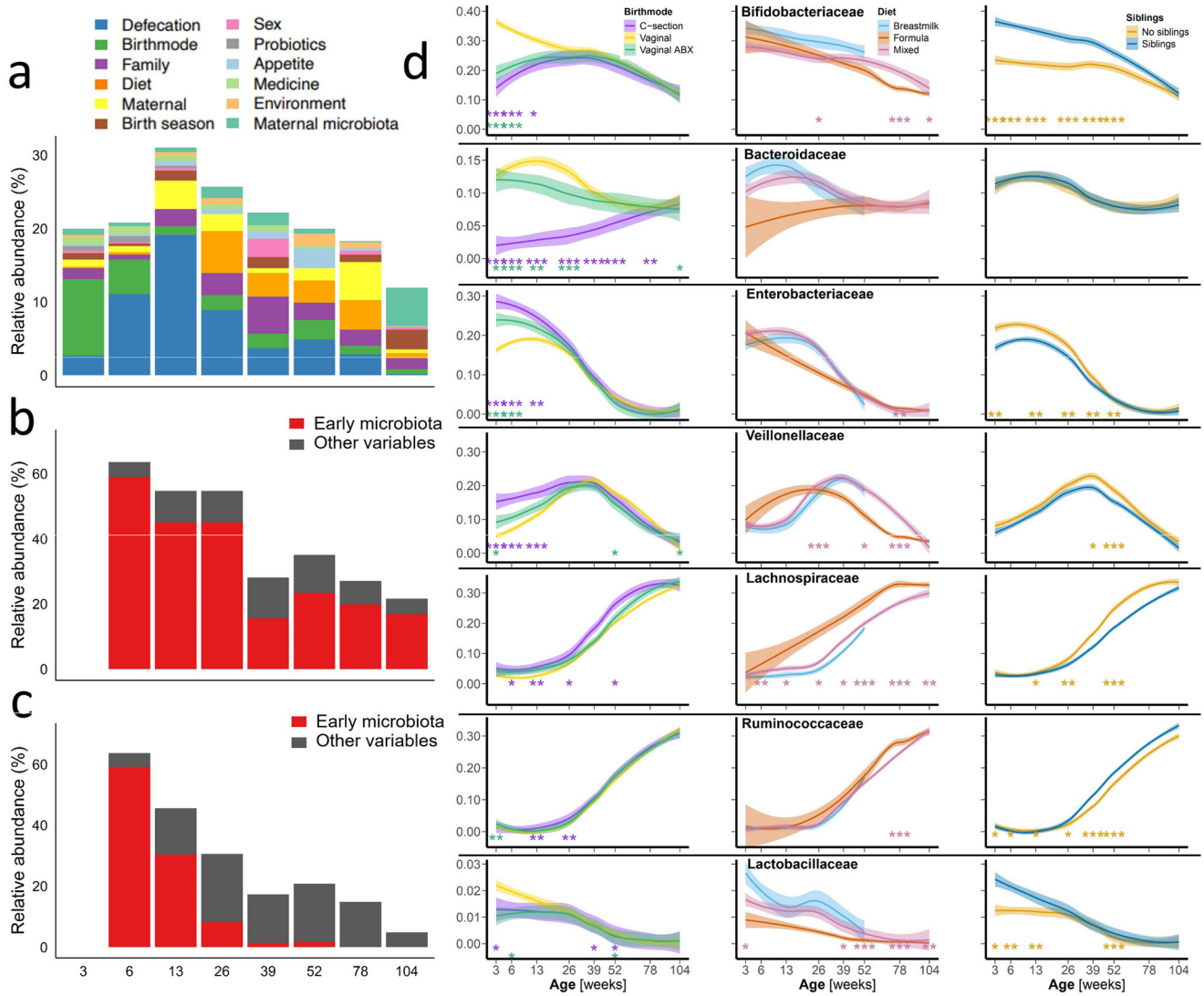

Fig. 2 | Impact of exposures and priority effects. a Variance in principal component (PC) scores partitioned to different types of early exposures at each time point according to multivariate regression. **b** Variance in PC scores attributable to early exposures (Other variables) and the previous time points' PC scores. **c** Variance in PC scores attributable to early exposures (Other variables) and the first time point's PC scores. **d** Associations between abundant microbial families, birth mode, breastfeeding, and siblings. Comparisons of Birthmode (C-section and Vaginal ABX to Vaginal), Diet (Breastmilk and Mixed to Formula) and Siblings (No Siblings to Siblings) are represented with *P*-values denoted by colour-coded asterisks at the bottom of the panel based on generalised linear models: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. The data are presented as mean values +/− SEM.

negatively correlated with the next time point's relative abundances of *Clostridiaceae, Enterobactericeae*, and *Ruminococcaceae*, and positively with members of *Bacilli, Actinobacteria* and *Bacteroidia* (Supplementary Fig. 4).

Microbiota community types in infants

To identify the main microbiota community types in the infants, we clustered the infant samples at the genus level using K-means clustering and log-pearson distance, identifying four community types (Fig. 3a). We tested additional distance metrics alongside the log-pearson derived community types: Bray-Curtis, Jaccard and Aitchinson (Supplementary Fig. 5a–c). The same community types are replicated irrespective of the distance metric and see an overall similarity between the metrics. Community types 1 and 2 (C1 and C2) characterized the first 26 weeks' microbiota, C3 was common at 39–52 weeks and C4 thereafter. C1 was dominated by *Bifidobacterium* (39.2% relative abundance) and *Bacteroides* (12.8%), together with other members of Actinobacteria and Bacteroidia covering over 50% of the relative abundance on average (Fig. 1b). Community type 2 (C2) was nearly devoid of bifidobacteria (4.8%), having a high relative abundance of *Clostridiaceae* (13.4%), and *Enterobacteriaceae* (25.7%). In community type 3 (C3), *Bifidobacteriaceae* (27.3%), *Lachnospiraceae* (18.5%), and *Veillonellaceae* (20.1%) were the dominant families, and community type 4 (C4) was dominated by *Lachnospiraceae* (30.0%) and *Ruminococcaceae* (30.0%) (Supplementary Data 5).

Microbial richness varied significantly between community types, being highest in C4 and lowest in C2 (Fig. 3c), and generally showing an increasing association with infant age. The relative abundance of potential pathobionts was the most abundant in C2 (Fig. 3d). Stool colour was significantly different between the different community types ($p < 0.0001$, c^2 test, Supplementary Fig. 6), with C1 and C2 being more likely to have yellow and green colour stool while C3 and C4 were observed mostly in brown stool, representing the change in stool composition when the amount of solid foods increases in infants' diet after 6 months. In the early months (C1 and C2) green stool was more likely to occur in C2 ($p < 0.0001$).

To explain the infants' community type, we used recursive partitioning. A model with 4 variables explained 56% of the variation in

https://doi.org/10.1038/s41467-024-52561-6

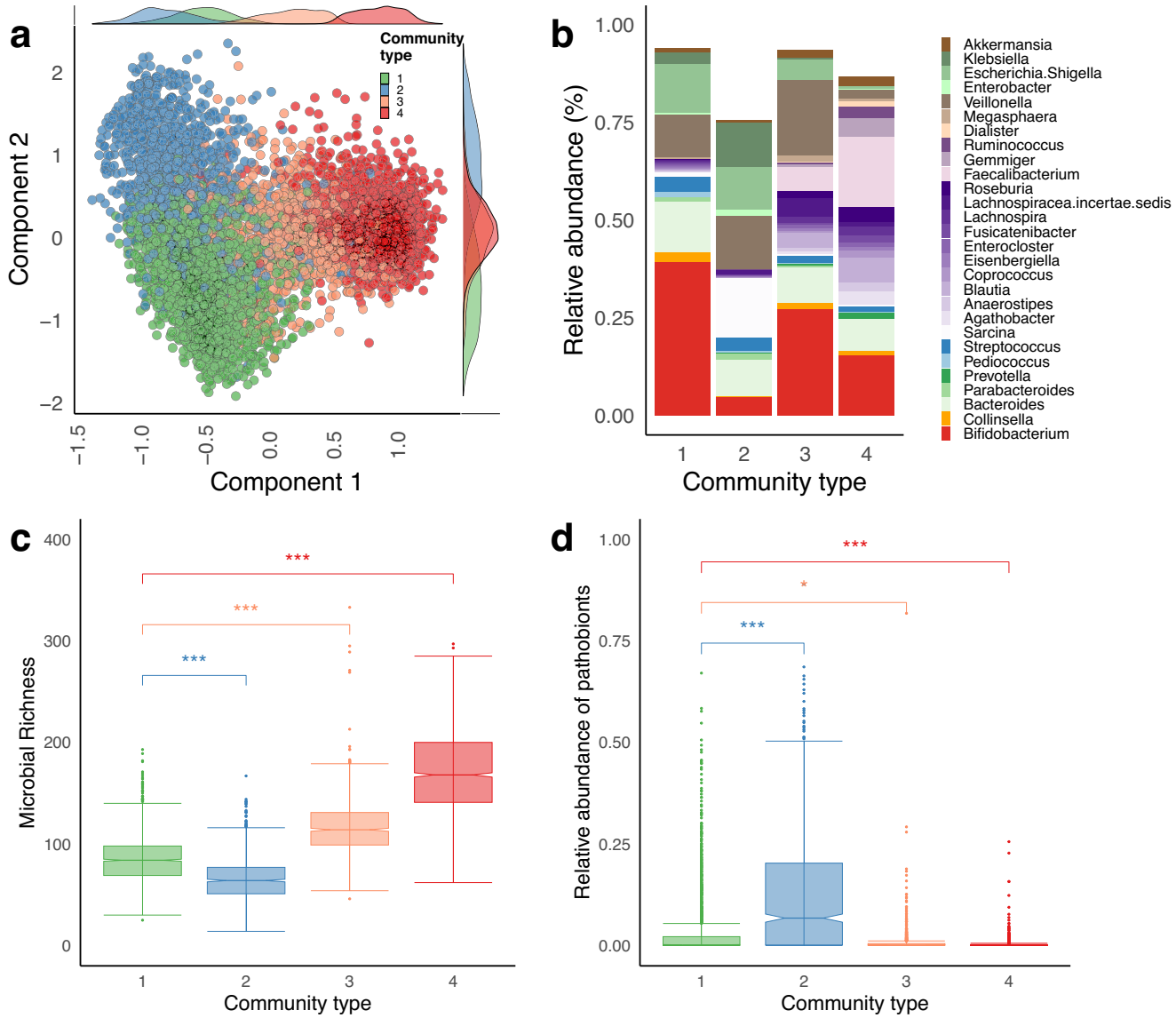

Fig. 3 | Microbiota community types. a Principal coordinates (PC) analysis on Pearson correlation distances of log-transformed relative abundances of microbial genera. Circle colour represents community type: C1: 2033 (green), C2: 1016 (blue), C3: 1100 (orange), and C4: 1932 (red). Histograms show distribution by community type for corresponding PC. **b** Average genus-level microbiota composition by community type ("C") in infants. Unallocated percentages are due to additional genera not shown. **c** Microbial richness (number of species). Significance of the difference to C1 is indicated by the asterisks:***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. All p-values < 2.26e-16. **d** Relative abundance of pathobionts by community type. Significance of the difference to C1 is indicated by the asterisks: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$ ($p < 2.26e-16$ (C1-C2), $p = 0.023$(C1-C3), $p < 2.2e-16$ (C1-C4)). Box plots show the median as centre line and interquantile range, and whiskers show the minimum and maximum quantile range Comparisions for (**c**, **d**) were done using a two-sided Wilcox rank sum test, and no adjustments were made for multiple testing.

community types (Fig. 4a). Age was the most important explanatory variable. The early communities, C1 and C2, were dependent on birth mode, siblings, and bifidobacteria-containing probiotics (ever taken prior to the sample), C1 being typical in the first 6 months of life in the vaginally born infants that had not been exposed to intrapartum antibiotics. Before the age of 6 months, CS-born infants were typically in C2, but by 6 months those that had siblings had often transitioned to C1. The V-ABX infants' samples were also classified into C2 before the age of 6 months, unless they had received bifidobacteria-containing probiotics or had siblings, which facilitated their transition to C1. At the age of 9 months, most infants were in C3. At 12 months, having siblings promoted early transition to C4. After 12 months, most infants were in C4 (Fig. 4a).

We tested the associations between community types at each time point and health outcomes at 2 and 5 years and discovered that C2 was associated with increased risk of undesirable health outcomes, especially allergic diseases (Fig. 4b, Supplementary Fig. 6, Supplementary Data 6). Children in C3 before the age of 6 months had an increased risk of allergic diseases and height-for-age Z-score < −1 sd at 5 years. Early transition to C4 (12 months) was associated with height-for-age Z-score < −1 sd at 2 years, but at 2 years, C4 was negatively associated with concurrent asthma diagnosis. At 12 months, C1 was associated with having had gastrointestinal infections.

Developmental trajectories

As the microbiota development was found to follow a consistent and predictable pattern in the individual infants, we utilized group-based trajectory modelling of the microbiota cluster scores to identify different patterns of microbiota development. Five distinct developmental trajectories were identified, with differences that mostly

https://doi.org/10.1038/s41467-024-52561-6

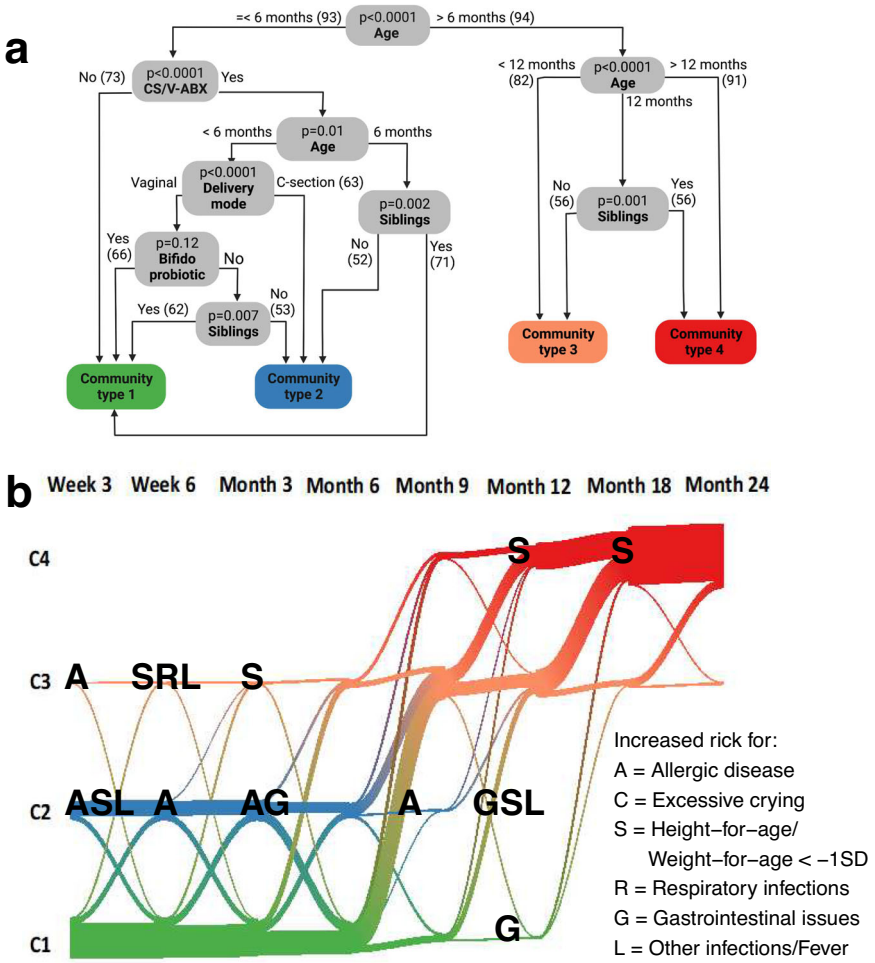

Fig. 4 | Community type determinants and health associations. a Partitioning tree of the determinants of microbiota community types. *P*-values from χ^2 test for the split and percentages (shown in brackets) of infants following the rule are shown. **b** Transitions between community types over time and the associations with health outcomes by 5 years ($p < 0.05$) at different ages using a negative generalized linear model (Supplementary Fig. 5, Supplementary Data 6). Lines representing < 5 infants are omitted. Adjustments are made for multiple comparisions.

manifested over the first 6 months. Trajectory 1 (T1) was the most common one ($N = 388$, 47%), characterized by stable C1 membership in the first 6 months, transition to C3 by 9 months, and to C4 by 12–18 months (Fig. 5a). These infants had a high initial relative abundance of *Bifidobacterium*, which declined gradually, being replaced initially by *Veillonella* and then by *Faecalibacterium* and members of *Lachnospiraceae* (Supplementary Fig. 8a). Infants in trajectory 2 (T2, $N = 95$, 11%) were initially in C1, but moved to C2 before transitioning to C3 (Fig. 5b), showing a rapid decline in *Bifidobacterium* and a transient increase in *Clostridium* and *Klebsiella* (Fig. 5b, Supplementary Fig. 8b). Infants in T3 ($N = 78$, 9%) began in C1 but oscillated repeatedly between C1 and C2 in the first 6 months (Fig. 5c, Supplementary Fig. 8c). The reverse pattern of T3 was represented by T4 ($N = 151$, 18%), where infants that started in C2 oscillated between C1and C2 in the first 6 months, showing a peak of *Bifidobacterium* at 6–9 months (Fig. 5d, Supplementary Fig. 8d). Infants in T5 ($N = 116$, 14%) were consistently in C2 throughout the first 6 months with a high relative abundance of *Clostridium* and *Klebsiella* (Fig. 5e, Supplementary Fig. 8e). The intraindividual similarity over time for the 5 trajectories (Supplementary Fig. 9) shows that microbiota development in general was the most rapid at 6–9 months, with infants in T3 showing the highest volatility at 3–12 weeks, and those in T2 at 9 months. Infants in T1 had generally the most stable microbiota compositions.

We compared the trajectories to our earlier data on average infant gut microbiota compositions around the world at class/phylum level, collected from 30 studies and 5732 infants in the first two years of life[8]. T1 resembled most closely the global normal development pattern, while T2 and T3 resembled the average compromised pattern (Fig. 5f). These results indicate that the trajectories that we identified here in a Finnish cohort can be recapitulated, at least partly, in other cohorts.

Associations between background factors and trajectory membership were assessed using the χ^2 test (Fig. 5g). Trajectories 1–3 were associated with vaginal delivery, while T4 and T5 were associated with CS birth and antibiotic prophylaxis during vaginal birth. In contrast to the other trajectories, T1 was associated with having siblings, living in a single-family house, and no formula feeding in the first 12 months. The transition to C2 exhibited by infants in T2 may have been promoted by formula feeding or lack of siblings, and potentially reflected in symptoms, as these infants were more likely than others to have received probiotics. The only identifiable reasons for the fluctuations between C1 and C2 in T3 was the lack of siblings in T3 and the possibly lower socioeconomic status indicated by housing type. Perhaps counterintuitively, infants in T3 were less likely than others to have received antibiotics in the first 3 months. The spontaneous microbiota correction in T4 may have been driven by breastfeeding, or other factors related to higher socioeconomic status.

The trajectories were tested for associations with infant health and wellbeing (Supplementary Data 7, FDR adusted *p*-values Supplementary Data 8) over the follow-up from birth to 2 ($N = 984$) and 5 years ($N = 496$) of age (Fig. 5h), and fever and several infection types

https://doi.org/10.1038/s41467-024-52561-6

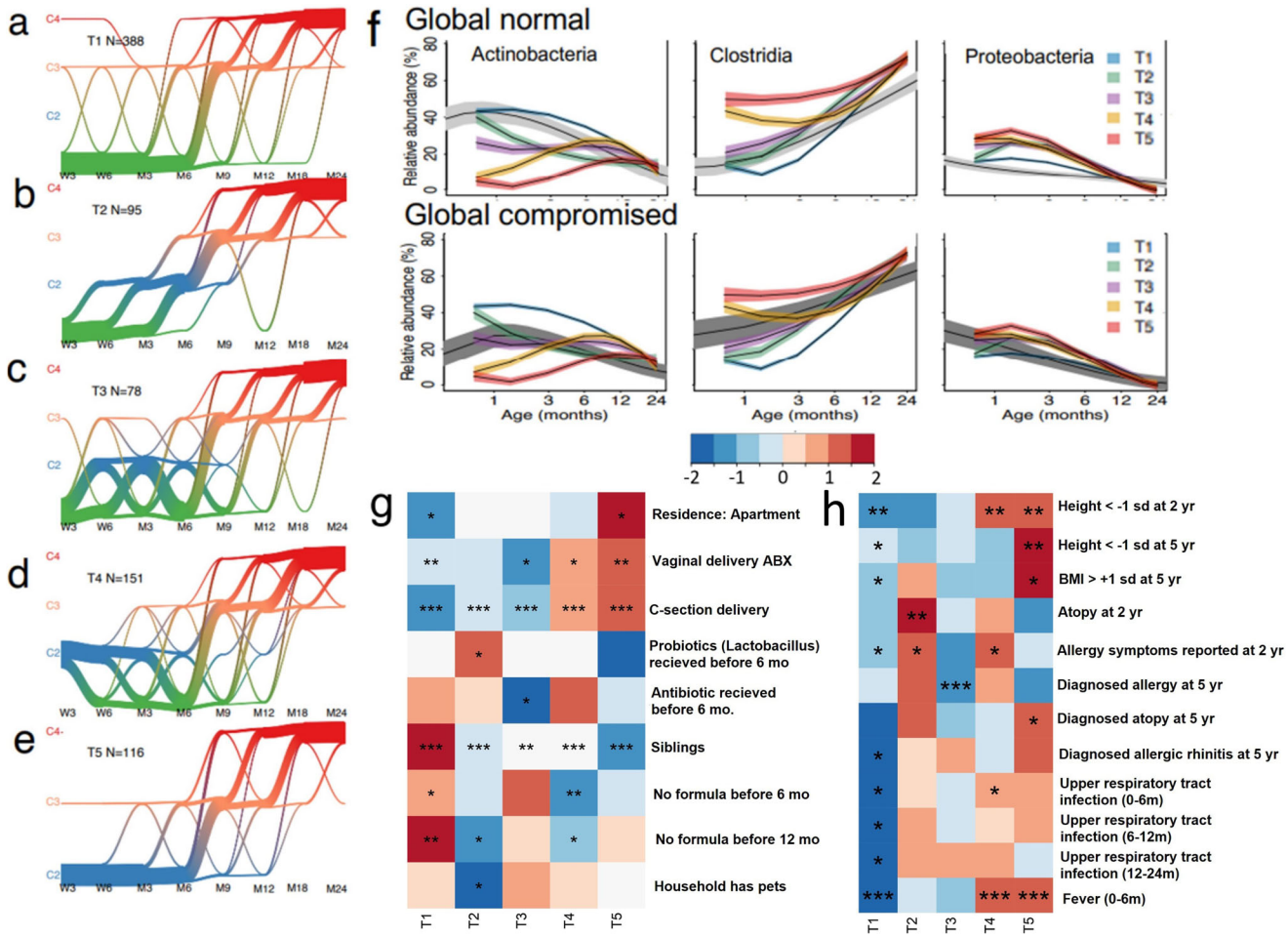

Fig. 5 | Microbiota development trajectories, determinants, and associations with health outcomes. Transitions between microbiota community types (C1:green, C2:blue, C3: orange, C4: red) by trajectory (**a**–**e**). Comparison between the trajectories (T1: blue, T2:green T3: violet, T4: yellow, T5: red) and average gut microbiota composition in global reference data representing 30 studies from 5732 infants divided into the normal (born vaginally, not exposed to antibiotics, and breastfed), and the compromised group (C-section born, antibiotic exposed or not breastfed)[8]. The trend line amd 95% confidence interval (shaded area) are obtained from a regression model with the third degree polynomial of log-transformed age, weighted by the size of the cohort. **f** The global data are cross-sectional averages, and thus represent very broad general patterns. Associations between trajectories and infant exposures, based on c^2 test on contingency tables (**g**), and between trajectories and health outcomes using logistic regression model (**h**). *P*-values denoted by asterisks at the bottom of each panel: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. Adjustments were made for multiple comparisons for **g**, **h**.

which were recorded at 0–3, 0–6, 6–12 and 12–24 months. After adjusting for parental allergies and education level, maternal BMI, paternal smoking, maternal smoking prior to pregnancy, gestational diabetes, pregnancy weight gain, infant sex, pets, siblings, birthmode and the sequencing run ID used at each time point, T1 was negatively associated with the following: reduced risk of reported allergy symptoms in the first 2 years, upper respiratory infections in the first 2 years, fever reported in 0–6 months, doctor-diagnosed allergic rhinitis at 5 years, ISO-BMI Z-scores[32] >1 standard deviation (>+1 SD) at 5 years, and height-for-age Z-scores less than −1 standard deviation (<−1 SD) at 2 and 5 years. T2 was associated with an increased risk of atopy at 2 years, and parent-reported allergy symptoms during the first 2 years ($p < 0.05$, $p < 0.01$, and $p < 0.05$ respectively). T3 was associated with a decrease rick of parent reported allergy symptoms at 2 year ($p < 0.001$). T4 and T5 were both associated with height-for-age < −1 SD at 2 years, and upper respiratory infections and fever between 0–6 months ($p < 0.01$; $p < 0.05$; $p < 0.001$). However, possibly due to the microbiota correction in T4, these infants did not have the increased risk for altered growth or diagnosed allergic rhinitis at 5 years that were observed in T5 ($p < 0.05$). T5 was additionally associated with increased rick of being diagnosed with atopy at 5 year ($p < 0.05$). Trajectory membership was more strongly associated with health outcomes than community type membership at any given time point (Supplementary Data 9), indicating that longitudinal analysis of development is more informative than single time points.

As an alternative way to represent microbiota development, we constructed a microbiota maturity index based on age-associated microbes. We found only minimal associations between the maturity index and health outcomes at different ages, mostly regarding growth (Supplementary Fig. 10a). We then compared the index to the trajectories, which did not greatly differ (Supplementary Fig. 10b), indicating that the maturity index was an insufficient representation of microbiota development. In the total data, we identified a set of bacterial genera associated with age (Supplementary Fig. 10c–g). Overall, the taxa displayed similar patterns across the trajectories, being broadly divided into early (members of Actinobacteria, Bacteroidia, Enterobacteria, Negativicutes, Bacilli) and late infancy (mainly Clostridia) groups. However, certain key groups such as *Bifidobacterium* and *Bacteroides* showed different temporal patterns in the different trajectories (Supplementary Fig. 10c–g). Species level community type and developmental trajectories were tested in addition to the genus level. The PC space and community type are highly similar to genus level (Supplementary Fig. 11a). Using the same criteria for trajectory creation we show that the background factor-trajectory associations

are similar as the genus level, with minor differences (Supplementary Fig. 11b, c).

Microbiota wellbeing index

Due to the inability of the maturity index to differentiate between different developmental trajectories and to capture various health associations, an alternative method to characterise microbiota wellbeing was devised. Because T1 was the most common pattern, associated with vaginal birth and positive health outcomes, and most representative of the normal gut microbiota development globally, we took this pattern to represent the natural undisturbed gut microbiota development ("eubiosis"). To identify microbes associated with natural gut microbiota development, we formed a reference group of infants in T1 that did not have diagnosed allergic diseases, allergic symptoms or atypical growth (absolute WHO Z-scores > 2 SD) during the first 5 years of life ($N = 198$). We used logistic regression to identify microbes predictive of membership in the reference group at different ages. The estimates from the model for the predictive bacteria were used as microbiota wellbeing influence scores, positive scores indicating that the microbe was associated with the reference. The strongest overall positive microbiota wellbeing association was seen in *Bifidobacterium* and *Bacteroides*, which were consistently indicative of the reference gut microbiota (Supplementary Fig. 6a). Most taxa showed an age-dependent association, either becoming increasingly positive with age (*Eisenbergiella, Oscillibacter, Parabacteroides, Anaerostipes, Streptococcus*), increasingly negative with age (*Lachnospira, Faecalicatena, Lacrimispora, Klebsiella, Sutterella*), showing a transient negative association (*Roseburia, Faecalibacterium*), or a transient positive association (*Citrobacter, Blautia, Gemmiger, Hungatella*). The amount of variance of the index explained by each microbe by time point further elucidates the age-dependency (Supplementary Fig. 12). The indicator microbes were individually assessed against the health outcomes at each time point (Supplementary fig. 13), verifying that these microbes were associated with health and wellbeing in a consistent manner.

The relative abundances of the indicator microbes were used to create a microbiota wellbeing index (MWI), representing the microbiota-based estimated likelihood of belonging to the reference group. The MWI was significantly lower in infants with an allergic disease or growth differences (Fig. 6b) and in more detailed analysis the MWI was associated with several different types of health outcomes from allergic diseases to growth at both 2 and 5 years of age (Supplementary Data 7), and the incidence of infections (Fig. 6c).

MWI was reduced in CS ($p < 0.0001$) and V-ABX exposed ($p < 0.0001$ at 3–6 weeks, $p < 0.0513$-104 weeks) infants throughout the first 2 years (Fig. 6d). Having siblings increased MWI in the vaginally born not antibiotic exposed ($p < 0.0001$ at 3 weeks, $p = 0.003$ at 26 weeks, $p = 0.008$ at 104 weeks), but in the V-ABX infants the effect was not observed until 26 weeks ($p = 0.003$), and in CS infants not until 78 weeks ($p = 0.02$). When analysed together, the impact of siblings on the CS/V-ABX infants was significant but weak at 3 weeks, strengthened at 26 weeks and remained significant until 104 weeks ($p = 0.01$,$p = 0.0003$, $p = 0.007$, respectively). Exclusive breastfeeding created a modest increase on the WMI in the vaginally born non-antibiotic exposed infants at 13 weeks ($p = 0.034$), but the main impact of breastfeeding was observed at the time of solid foods' introduction (26 weeks), when those that were no longer breastfed experienced a significant drop in MWI ($p = 0.001$ at 26 weeks, $p = 0.002$ at 39 weeks, $p = 0.038$ at 52 weeks), while breastfed infants retained a high MWI. Breastfeeding was not associated with MWI in the C-section born infants, but modestly increased MWI in the V-ABX infants in the first 13 weeks ($p < 0.05$). When analysed together, breastfeeding increased MWI in the CS/V-ABX infants in the first 13 weeks ($p < 0.01$).

Discussion

Utilising a longitudinal cohort of 984 term-born infants followed from birth to the age of 2–5 years, we delineated the general patterns and determinants of gut microbiota development in the first 2 years of life and created a gut microbiota wellbeing index, that was associated with the health and wellbeing of the child. Our results show that early-life gut microbiota development is highly predictable, dependent on infant exposures and thus malleable, and predictive of physiological development of the host. These results highlight the importance of biotic interactions in the gut on both host and microbiota development.

We identified four community types in infants, mainly differentiated by the relative abundance of *Bifidobacteriaceae, Enterobacteriaceae, Clostridiaceae, Veillonellaceae, Lachnospiraceae*, and *Ruminococcaceae*, and based on the dynamics of the community types, five major developmental trajectories. Our results recapitulate and extend the results from previous studies with lower numbers of infants, shorter time courses and shallower sampling addressing infant microbiota clusters[26] and trajectories[20,25]. We show that the infant microbiota development is not stochastic but follows specific temporal development trajectories where the initial composition has a profound long-term impact. The community types at different ages and the developmental trajectories were associated with host health outcomes at 2 and 5 years, supporting the notion that the overall gut microbiota balance may influence the long-term physiological development of infants. Especially in compositional data, where the relative abundances of different taxa are interdependent, the approach of identifying microbial community types and trajectories linked with host outcomes may be more meaningful than analyzing individual taxa as separate entities. By utilizing different distance metrics with the log-pearson derived community types, we show that similar patterns exist regardless of the metric used.

Capitalizing on the large longitudinally monitored HELMi cohort, we characterized gut microbiota development in infants in relation to health. We discovered that there was one health-associated developmental trajectory, and several trajectories associated with different types of undesired health outcomes and symptoms. We thus defined the healthy microbiota development, eubiosis, and several deviating, or dysbiotic, patterns in the infant gut. Importantly, not all infants follow the same trajectory and thus a general maturity index may not capture the full spectrum of developmental differences. The health-associated trajectory, T1, was the most common trajectory and associated with vaginal birth without antibiotic exposure, breastfeeding and having siblings—all factors known to contribute positively to healthy gut microbiota in infants[33,34]. The development of the most abundant taxa followed a similar pattern in T1 as we have previously observed globally in vaginally born breastfed infants[8]. This indicates that this trajectory may indeed represent the natural gut microbiota development in human infants beyond this cohort. Notably, a recent study[10] suggested that infants from Northern Europe exhibit a unique microbiota composition with high abundance of *Bacteroides* spp. This may reflect the uniqueness of the particular cohort or may be a technical artefact emerging from DNA extraction or PCR bias, and not present in our data. It is however possible that our results do not generalise to populations with very different genetic backgrounds or lifestyle patterns.

Each trajectory was associated with a unique health profile (Supplementary Data 9), suggesting that the sequence of microbiota development in the first 6 months, rather than simply microbiota composition at a given time point, influences infant development. Spontaneous correction of gut microbiota by 6 months, as was observed in trajectory 4, ameliorated some, but not all, of the health risks associated with the earlier dysbiotic composition, indicating that the time window for gut microbiota correction and healthy immune imprinting in infants may close before the age of 6 months. While the health outcomes were primarily recorded at or after 2 years of age,

https://doi.org/10.1038/s41467-024-52561-6

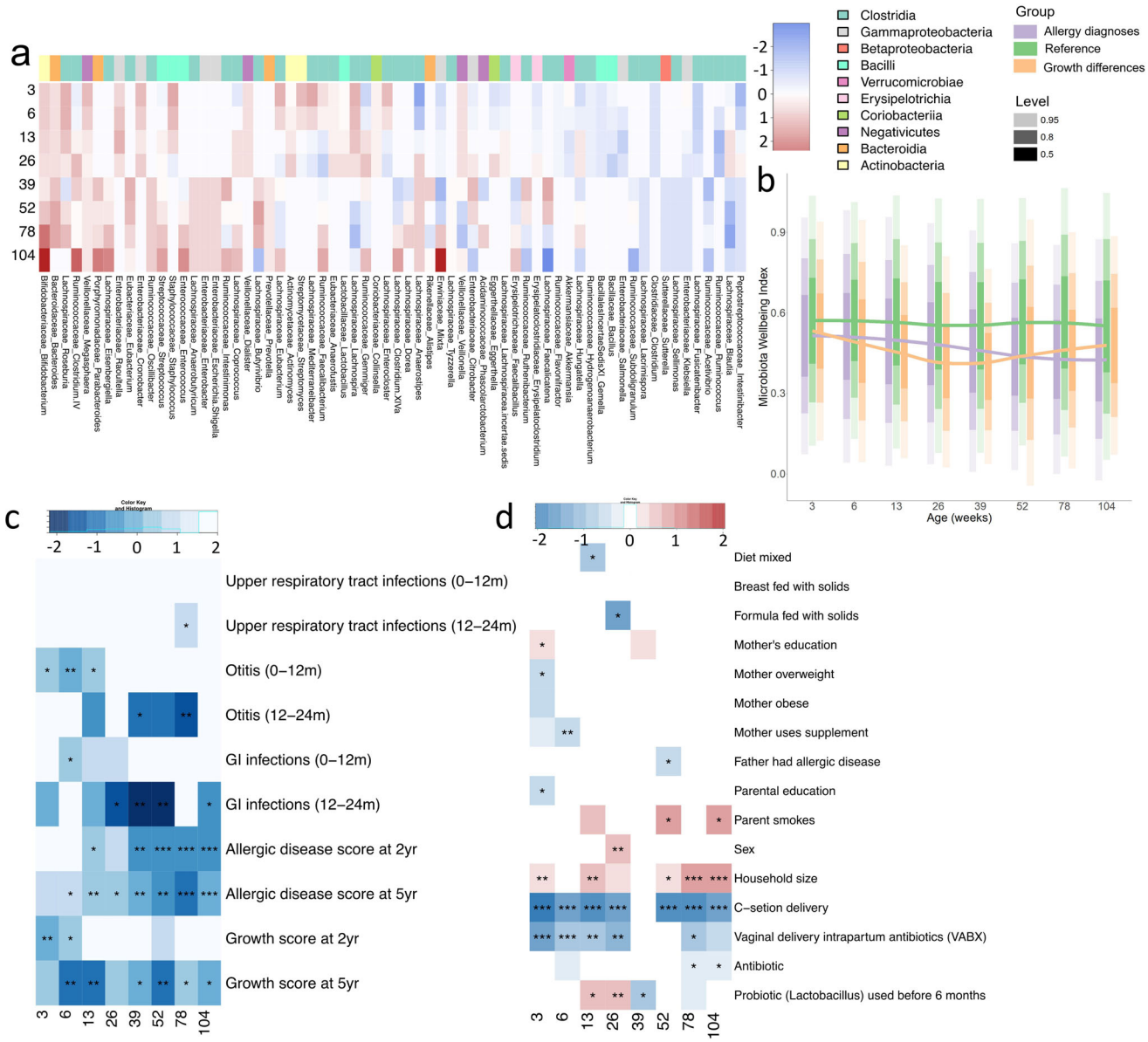

Fig. 6 | Microbiota wellbeing index (MWI). a Wellbeing index and microbial associations, measured as estimates from a logistic regression model predicting membership to the reference group. Top bar colours represent the bacterial class. **b** Wellbeing index values over time for Reference (green;) defined as children with no diagnosed allergic diseases or growth below or above the 2 sd during the first 5 years of life in T1, Growth differences (orange) defined as those children with weight below/above the 2 sd and height <2 sd during the first 5 years of life, and Allergic diseases (purple) defined as those children with allergy, atopy, rhinitis or asthma during the first 5 years of life, showing the means (lines) and the percentage (0.5,0.8,0.95) of the population within the scale. **c** Area under the curve (AUC) values from logistic regression models predicting health outcomes with MWI measured at different ages: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $\cdot\ p < 0.1$. Only prospective associations are included. **d** Associations between MWI and infant exposures. No adjustments were made for multiple comparisions.

fever and infections were recorded earlier.These could themselves alter the gut microbiota and thus reverse causation cannot be ruled out. High body temperatures has been seen to influence levels of bile acid in serum and intestine in a mouse study[35], while a systematic review showed an increase *Enterococcus*, and a decrease in *Firmicutes*, *Lachnospiraceae*, *Ruminococcaceae*, and *Ruminococcus* in patients with respiratory infections, although the direction of the association is not clear. The gut-lung axis[36] suggests the gut microbiota may have an effect on the propensity for respiratory tract infections[37,38]. Healthy gut microbiota development in infants was characterized by high and stable relative abundance of *Bifidobacterium* in the first year, and low relative abundance of pathobionts. Indeed, the *Bifidobacterium* genus emerged as a key indicator of gut microbiota wellbeing throughout the first two years of life. It appears likely that *Bifidobacterium spp.* have both direct beneficial effects on the host[39] and a strong guiding influence on the gut microbiota, as they were the most important early driver of gut microbiota development, together with *Bacteroides*. We have previously shown that supplementation of C-section born infants with infant gut adapted *Bifidobacterium spp.* can ameliorate the C-section-induced microbiota imbalance[40], and reduce the incidence of allergic disease[41], experimentally validating the important role of *Bifidobacterium spp.* in infants. Indeed, *Bifidobacterium spp.* are one of the most clearly health-associated gut bacteria in infants[42,43], where their early abundance predicts positive later health outcomes[43,44], and are associated with beneficial effects in adults, as well[45–47]. There is much less information on the role of *Bacteroides* spp. in infant health, but they have been associated with positive effects on neurodevelopment[48].

A key trait that certain infant-adapted species of *Bifidobacterium* and *Bacteroides* is the human milk oligosaccharide (HMO) utilization[49]. Their ability to unlock substrates from HMOs likely critically influences the growth of other beneficial taxa[50–52] in addition to transforming HMOs into short-chain fatty acids (SCFAs) that may contribute to child nutrition and immune development, e.g., reducing the risk of atopic-related outcomes in childhood[53]. Indeed, we recently showed that infant gut communities with low *Bacteroides* and *Bifidobacterium* abundances have reduced relative abundance of carbohydrate degradation pathways[54]. Furthermore, *Bifidobacterium spp*. have inhibitory effects against taxa such as staphylococci and streptococci[55,56]. Our results support the important role of species of *Bifidobacterium* and *Bacteroides* spp. as keystone organisms in the infant gut, as we and others have suggested[57], regulating the overall composition through both facilitative and inhibitory effects. Notably, *Bifidobacterium* and *Bacteroides* are the two genera that are consistently depleted in CS born infants[8,58], as also shown in our data, because they are dependent on vertical transmission from the mother's gut at birth[5,6].

Beyond *Bifidobacterium* and *Bacteroides* spp., *Lactobacillus, Roseburia, Anaerobutyricum*, and *Eubacterium* spp. were among the top indicators of a healthy early-life gut microbiota. Many strains of lactic acid bacteria are used as probiotics and have been shown to have some beneficial effects on the gut microbiota of infants[59,60], partly due to their bacteriocin production that can inhibit the growth of pathobionts[61], and e.g., *Limosilactobacillus reuteri* has reportedly been effective in the treatment of colic[62]. *Eubacterium rectale* and *E. hallii* (or *Anaerobutyricum hallii*) have been identified as key butyrate producers in early life[60,63], that cross-feed with *Bifidobacterium* and *Akkermansia* spp[51,64]. In order to have a more general MWI that is more accessible we did not use alpha diversity or very uncommon taxa in the wellbeing index development as these indeed may vary between cohorts.To the best of our knowledge, this is the first study proposing a gut microbiota index associated to wellbeing and health in children, where the prospective setting allows addressing directionality. Analogous approaches have been taken in adults, and the indicator organisms that we identified were partly similar to those included in an adult Gut Microbiome Health Index[65–67]. Despite the compositional differences between the infant and adult microbiota, notable parallels in terms of dysbiosis indicators included *Flavonifractor, Blautia, Clostridium, Klebsiella*, and *Veillonella* species. These similarities support the generality of our results and suggest that they may apply to infants in other cohorts, as well. The genera *Clostridium* and *Klebsiella* contain many known pathogens[68]. Species of *Veillonella* produce H_2, that has been associated with colic in infants[69].

Our results show that the initial delay in the colonization of the gut by keystone HMO-utilising bacteria alters the long-term developmental trajectory of the gut microbiota due to their dominant ecological role. Indeed, priority effects emerged as the strongest influence on gut microbiota development, implicating ecological interactions between bacteria, determined by birth mode-induced priority effects, rather than postnatal external influences, as the major driver of gut microbiota dynamics in infants. The importance of initial community assembly and HMO metabolism on infant gut-associated bifidobacterial communities has been experimentally validated[50]. Our results highlight the importance of the initial inoculum that an infant receives at birth from the mother, even though maternal gut microbiota composition does not strongly influence the neonatal gut microbiota; specific maternal microbes are selected by the infant's diet and gut environment - hence the maternal composition is not maintained until it begins to reshape when the infant transitions to solid foods. The predictability of microbiota development analysis only included the first two PCs and thus some variation is absent from the model.

Notably, all the taxa reduced in CS-born infants were indicators of a healthy gut microbiota. Intrapartum antibiotic exposure in vaginal deliveries reduced mainly the Gram-positive bacteria but not the Gram-negative ones, i.e., affected only a part of the microbiota, explaining its milder impact on MWI. While the overall microbiota composition usually normalises by 12 months after CS birth[58,70], the MWI was still significantly reduced in CS-born and antibiotic exposed infants at 2 years, showing that the health-associated microbiota recovers slower than the overall composition. However, the recovery was accelerated by exclusive breastfeeding and in those with siblings. While siblings emerged as one the most influential factors supporting a healthy gut microbiota development, the early impact of siblings appeared to arise through maternal transmission of infant-adapted *Bifidobacterium* spp., as the effect was delayed in C-section born infants[71]. Together with earlier results on the effect of family size on infant microbiota variation[71,72], our results highlight the metacommunity structure of human gut microbiota and the importance of vertical transfer of microbes at birth. Our results suggest that siblings begin to exert a direct influence on the gut microbiota of infants after the first 6 months, mainly by introducing *Ruminococcaceae* that are indicative of microbiota maturation. Interestingly, having older siblings has long been known to be protective against allergic diseases[72,73] This was originally attributed to lower levels of hygiene and increased transmission of infections, but our results indicate that the protective impact of siblings may be mediated by the increased transmission of beneficial gut microbes, rather than pathogens. A similar finding was recently shown in a large cohort of infants where sibling-induced microbiota maturation at 1 year was associated with lower risk of food allergy[73]. We have previously discovered that the number of siblings is positively associated with Proteobacteria and Firmicutes in dust microbiota of Finnish homes, suggesting increased intra-familial microbial transfer in homes with multiple children[73].

Our results show that exclusive breastfeeding is insufficient in restoring the gut microbiota of CS-born infants. This indicates that breastmilk is not a sufficient source of gut microbes to the infant but promotes the growth of microbes that an infant receives during vaginal birth. Bacteria with mucus-degrading capacity, such as *Akkermansia* and *Ruminococcus*, were associated with lack of breastfeeding in the first months, likely as an indication of a substrate-starved microbiota relying on host-derived mucus. Indeed, the gut microbiota of formula-fed infants has been shown to exhibit increased mucus-degrading activity linked to *Bacteroides* spp. and associated with weakened gut barrier[74]. Furthermore, infants who were weaned at the time of solid foods' introduction experienced a drop in MWI, suggesting that continued breastfeeding through the first year, as recommended by WHO, may help the gut microbiota adapt to the new diet.

Our study faces limitations as it focuses entirely on the bacteriome of the gut and does not consider the other gut microbiota including fungi, eukaryotes, phage and viruses.As all children from our cohort are located entirely in Finland, little variation in terms of diet, genetics, lifestyle, ethnicity or geographic location are taken into account in our study. While race and geography are possible sources of variation, its believed that human behaviour, such as birth mode and breastfeeding, have larger impacts on microbial colonization[75].

A specific pattern of health-associated infant gut microbiota development was characterized, hallmarked by high and gradually declining relative abundance of *Bifidobacterium*. A microbiota wellbeing index based on the natural developmental trajectory was consistently and significantly reduced in infants that had current or later health problems, indicating that such an algorithm can be used to identify infants at risk of developing allergic diseases or symptoms and highlighting the importance of gut microbiota to overall health and wellbeing of children. We identified Bifidobacterium and Bacteroides as key taxonomic groups in the infant gut, affecting the community assembly and host physiological development. Ongoing research is focused on improving the understanding of allergy development in children with regards to gut microbiota development.

Methods

This study has received written informed consent from parents/guardians for use of samples and data from the children used in this study. The study was approved by the ethical committee of The Hospital District of Helsinki and Uusimaa, Finland (263/13/03/03 2015) and performed in accordance with the principles of the Helsinki Declaration.

Study population

Our observation stuy uses the HELMi cohort[31], a longitudinal birth cohort from the Helsinki metropolitan area, Finland. In total 1055 infants and their parents were enroled in during the recruitment period (February 2016–March 2018). Only healthy, term, singleton infants born on gestational weeks 37–42 with birth weight exceeding 2.5 kg were included in the cohort[31]. Stool samples were collected from infants at ages 3 and 6 weeks, and 3, 9, 12, 18 and 24 months, and from parents near the delivery. Stool samples were collected and frozen (−20°) at home by the parents[31]. Samples were later transferred to −80 °C storage at the laboratory by the parents. We utilized the samples from 967 infants, 768 mothers, and 515 fathers, comprising 7211 faecal samples. Questionnaire data on lifestyle, environmental exposures, and the child's diet, health and wellbeing were collected with online questionnaires at different time intervals from weekly to one-time questionnaires until 2 years, as previously described in detail (cohort profile). In addition, we used follow-up data for 496 children at 4–5 years, which included information about the health and wellbeing of the children. Families with infants with a diagnosis of a serious long-term illness (other than allergic diseases) by 2 years were excluded ($N = 31$). Both biological and non-biological fathers ($N = 6$) were included and are referred to as fathers, however variables on health and disease history were linked to the biological father. Socioeconomic background variables (*e.g.*, level of education) addressed the acting guardian. Supplementary Data 1 lists the variables used in the study.

Child health, development, and well-being Health-related data (Supplementary Data 7) used in this study were collected from the parent-filled questionnaires and consisted of two types of variables:

(a) Doctor's diagnosis on allergic diseases at 2 years, and/or diagnosis on atopic dermatitis, allergic rhinitis and/or asthma at 5 years; phenotypic allergy i.e., symptoms of above-mentioned allergic diseases at 2 and/or 5 years of age, based on the International Study of Asthma and Allergies in Childhood (ISAAC) questionnaire[76]; age-adjusted height or weight > 2 or <−2 SD at 2 and/or 5 years based on nurse's measurements in postnatal care visits and transformed to age-dependent WHO z-scores[32], the occurrence of lower and upper respiratory infections, gastrointestinal (GI) infections, ear infections, pox infections, or fever. Infections which didn't fall into one of these categories were few ($N < 5$) and were excluded.

(b) Normal variations and/or parents' subjective assessment of child development or well-being was addressed using the following variables: age-adjusted height or weight > 1 or <−1 SD at 2 and/or 5 years based on nurse's measurements in postnatal care visits and transformed to age-dependent WHO z-scores[32], the lower median of parents' assessment of child's general health by a 0–100 mm visual analogue scale (VAS), collected every 3 months until age of 2; gastrointestinal function (defecation rate, parent-perceived stomach pain intensity and flatulence) until 2 years. Crying was addressed in the neonatal phase, grouping infants to upper and lower median based on daily crying time, averaged for weekly records by the age of 3 months.

Microbiota analysis

Bacterial DNA was extracted from the faecal samples using a previously described bead beating method[77] (Ambion MagMAX™ Total Nucleic Acid Isolation Kit (Life Technologies)) and KingFisherTM Flex-automated purification system (ThermoFisher Scientific) as previously described[71]. The 16S rRNA gene amplicon sequencing was performed using Illumina MiSeq and HiSeq platforms for V3-V4 (primers 341 F/785 R)[71] at the Functional Genomics Unit and Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. The sequencing reads were processed using R package mare[78], which relies on USEARCH[79] for quality filtering, chimera detection, and taxonomic annotation. Forward reads (V3), truncated to 150 bases, were used[80]. Reads occurring <50 times were excluded as potentially erroneous. The taxonomic annotation was performed using USEARCH[79] by mapping the reads to the Ribosomal Database Project taxonomy database version 18[81], restricted to known gut-associated taxa. Taxonomic annotations were verified using RDP classifier and in cases of disagreement, the Blast annotation was used. Potential contaminants were filtered by removing reads appearing in negative controls (PCR or extraction blanks) in corresponding numbers from all samples. The impact of the technical variation on this dataset is thoroughly explored in Jokela, et al.[82]. The sequencing depth cut-off was chosen to be 3000 reads after QC to samples collected at 3 months or before and to 5000 paired reads for the remaining samples based on species richness to sequencing depth evaluations. Richness was calculated as the number of species-level taxonomic groups in the sample. Lower sequencing depth and cross-platform sequencing have been shown to not adversely affect biological results.

Statistical analysis

The statistical analysis was performed using R version 4.2.2. The principal coordinates analysis was performed on Pearson correlation distances of log-transformed relative abundances of microbial genera using the vegan package[83]. K-means clustering was performed with the clustering package[84] using the log-Pearson distance. Using the gap statistic, we saw four optimal clusters: Cluster 1 ($N = 2033$), Cluster 2 ($N = 1034$), Cluster 3 ($N = 1100$), Cluster 4 ($N = 1932$). For the trajectory analysis, we selected infants that had at least one sample from 3–6 weeks, one sample from 3–6 months, and one sample from 1–2 years ($N = 828$ families) to ensure a proper temporal resolution of development. Trajectories were created using group-based multivariate trajectory modelling[85] on the clustering scores of each sample using the GBMT package[86]. The models used a 3rd degree polynomial, and the optimal number of groups based on Akaike information criterion (AIC) was 5: T1 ($N = 388$), T2 ($N = 95$), T3 ($N = 78$), T4 ($N = 151$), T5 ($N = 116$). Stool colour differences between community type were tested using the c^2 test.

Partitioning tree for cluster membership was created using the rpart library[87]. We utilized a complexity factor of 0.001 and a minimum split for each branch of 100. In creatin the partition tree we utilised 60 variables (Supplementary Data 1) ranging from technical, parental background, and infant related variables. Of these 60 variables, 5 remained in the partition tree. The relative abundance of potential pathobionts was calculated by summing the relative abundances of *Klebsiella*[88], *Fusobacterium*[89], *Sutterella*[90], *Bilophila*[90], *Salmonella*[90], *Haemophilus*[88], *Pseudomonas*[90], *Treponema*[91], *Campylobacter*[92], *Serratia*[93], *Vibrio*[90], *Proteus, Yersinia*[90], and *Neisseria*[90].

Associations between health outcomes (31 variables) and clusters and trajectories was done by performing a logistic regression model adjusting for mother's BMI, mother's and father's allergies, mother's and father's education level, infant's sex, presence of pets in the household, household size (number of people in the home), mother's weight gain during pregnancy, and gestational diabetes and run id. Technical variables were tested against both the baclground factors and health outcomes at each sampling age (Supplementary Data 10). Run id was found to be significantly associated with several variables and hence was included as a confounder For clusters analysis we performed the analysis separately by time point due to repeat measurements. We only tested associations for health outcomes before the outcome took place, except with regards to fever and infections which were recorderd at 0–3 months, 0–6 monhts, 6–12 months and

12–24 months. Associations between the trajectories with infant background variables (17 variables) were performed using the χ^2 test on contingency tables for each trajectory against the rest, as well as separately each trajectory against T1, and finally each trajectory against T5. T1 and T5 comparisons were performed as they represent the two extremes of the trajectories' patterns.

Comparisons of the relative abundance of the bacterial genera for categorical data were conducted using the GroupTest functions in mare[78]. This function selects an optimal model for each taxon using the lm function, glm.nb function from the MASS package[94] or the gls function from the nlme package[95]. The models were adjusted for probiotic and antibiotic usage prior to sample collection, number of siblings, birth mode, and diet (exclusive breastfeeding, mixed feeding, exclusive formula feeding, or one of these with solid foods). Supplementary Data 2, 3, 4, and 5 present the estimate, raw *P*-value, fold change, and the adjusted *P*-value (FDR) for the variables tested, and the model used for that taxon. *P*-value adjustment was done in order to reduce occurrence of false positives.The first two principal coordinates scores were used as response variables to assess the relative impact of different exposures on microbiota composition at different time points. We formed linear regression models with the exposures as explanatory variables and for each exposure variable calculated the average variance explained between components 1 and 2. Using these models and the microbiota composition at 3 weeks, we simulated the gut microbiota development in the infants with a full data set ($N = 98$) and calculated the correlation between the simulated and observed PC scores. The function adonis2 in the R package vegan was used to identify the most important bacterial taxa associated with the following time points' microbiota composition. We created the maturity index by selecting the genus level bacteria that occur in 40% of the samples, producing a list of 47 genera. The index was created fitting a negative binomial generalized linear model to predict age based on the microbiota, allowing for non-linear associations by including second-degree polynomials, using the glm.nb function from the MASS package[94].

Microbiota well-being index (MWI) was calculated by identifying a model that explained membership in the healthy reference cohort (healthy children), no diagnosed allergic diseases or growth below or above 2 standard deviation during the first 5 years of life in T1, $N = 198$, assigned a value of (1) as compared to healthy children in other trajectories (assigned 0.5) and children with variable health (with suspected or diagnosed allergic diseases or growth below or above 2 standard deviation during the first 5 years, assigned 0). Because the microbiota composition changed rapidly at 6–9 months, we fitted one model for the first 6 months and a second model for >6 months, including taxa that were present in >30% of the samples at the given age range. A logistic regression model with a second-degree polynomial age interaction with each taxon was fitted and AIC-based stepwise model reduction was used to arrive at a model for microbiota wellbeing. For each taxon, the model estimates were used as indices of microbiota wellbeing influence at each age. We used the prediction of the model (probability of a sample belonging to the healthy reference group) as the microbiota wellbeing index. The MWI was normally distributed and analyzed against birth mode, siblings, and diet by age using linear regression models. The MWI was analyzed against health outcomes using logistic regression. To identify the most important indicator microbes at each time point, we used a random forest model. Based on the node purity, we identified the top genera and calculated the amount of variance in the MWI explained by each of the genera by adding the top genera in a step-wise manner to a linear model, allowing for microbe-microbe interactions.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The HEMLi microbiome 16 s rRNA gene sequences in this study have been deposited in the European Nucleotide Archive (ENA) under accession code PRJEB55243, along with limited metadata (collection date, sex, age in weeks, geographic location, and sequencing method). Additional individual-level metadata, even pseudonymized, are sensitive and are protected by the GDPR and not publicly available. Reasonable data sharing requests based on data processing and material transfer agreements can be made to Anne Salonen, University of Helsinki, Finland. (anne.salonen@helsinki.fi)

Code availability

The R code used in the study is available at https://github.com/bhick001/Wellbeing-Index (https://doi.org/10.5281/zenodo.13359211)[96]

References

1. Sharon, G., Sampson, T. R., Geschwind, D. H. & Mazmanian, S. K. The central nervous system and the gut microbiome. *Cell* **167**, 915–932 (2016).
2. Martin, R. et al. Early life: gut microbiota and immune development in infancy. *Benef. Microbes* **4**, 367–382 (2010).
3. Nguyen, Q. P. et al. Associations between the gut microbiome and metabolome in early life. *BMC Microbiol.* **21**, 238 (2021).
4. Korpela, K. & de Vos, W. M. Infant gut microbiota restoration: state of the art. *Gut Microbes* **14**, 2118811 (2022).
5. Korpela, K. et al. Selective maternal seeding and environment shape the human gut microbiome. *Genome Res.* **28**, 561–568 (2018).
6. Korpela, K. et al. Maternal fecal microbiota transplantation in Cesarean-born infants rapidly restores normal gut microbial development: a proof-of-concept study. *Cell* **183**, 324–334.e5 (2020).
7. Senn, V. et al. Microbial colonization from the fetus to early childhood—a comprehensive review. *Front. Cell. Infect. Microbiol.* https://doi.org/10.3389/fcimb.2020.573735 (2020).
8. Korpela, K. & de Vos, W. M. Early life colonization of the human gut: microbes matter everywhere. *Curr. Opin. Microbiol.* **44**, 70–78 (2018).
9. Beller, L. et al. Successional stages in infant gut microbiota maturation. *mBio* **12**, e01857–01821 (2021).
10. Xiao, L., Wang, J., Zheng, J., Li, X. & Zhao, F. Deterministic transition of enterotypes shapes the infant gut microbiome at an early age. *Genome Biol.* **22**, 243 (2021).
11. Adlerberth, I. et al. Gut microbiota and development of atopic eczema in 3 European birth cohorts. *J. Allergy Clin. Immunol.* **120**, 343–350 (2007).
12. Marra, F. et al. Antibiotic use in children is associated with increased risk of asthma. *Pediatrics* **123**, 1003–1010 (2009).
13. Kostic, A. D. et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe.* **17**, 260–273 (2015).
14. Kalliomäki, M., Carmen Collado, M., Salminen, S. & Isolauri, E. Early differences in fecal microbiota composition in children may predict overweight. *Am. J. Clin. Nutr.* **87**, 534–538 (2008).
15. Conterno, L., Fava, F., Viola, R. & Tuohy, K. M. Obesity and the gut microbiota: does up-regulating colonic fermentation protect against obesity and metabolic disease? *Genes Nutr.* **6**, 241–260 (2011).
16. Korpela, K. et al. Childhood BMI in relation to microbiota in infancy and lifetime antibiotic use. *Microbiome* **5**, 26 (2017).
17. Bloom, S. M. et al. Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe.* **9**, 390–403 (2011).
18. Cho, I. et al. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* **488**, 621–626 (2012).
19. Gensollen, T. & Blumberg, R. S. Correlation between early-life regulation of the immune system by microbiota and allergy development. *J. Allergy Clin. Immunol.* **139**, 1084–1091 (2017).

20. Stewart, C. J. et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).
21. Martin, R. et al. Early-life events, including mode of delivery and type of feeding, siblings and gender, shape the developing gut microbiota. *PLoS ONE* **11**, e0158498 (2016).
22. Korpela, K. et al. Antibiotics in early life associate with specific gut microbiota signatures in a prospective longitudinal infant cohort. *Pediatr. Res.* **88**, 438–443 (2020).
23. Bäckhed, F. et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe.* **17**, 690–703 (2015).
24. Shao, Y. et al. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* **574**, 117–121 (2019).
25. Roswall, J. et al. Developmental trajectory of the healthy human gut microbiota during the first 5 years of life. *Cell Host Microbe.* **29**, 765–776.e763 (2021).
26. Galazzo, G. et al. Development of the microbiota and associations With birth mode, diet, and atopic disorders in a longitudinal analysis of stool samples, collected from infancy through early childhood. *Gastroenterology* **158**, 1584–1596 (2020).
27. Dai, D. L. Y. et al. Breastfeeding enrichment of B. longum subsp. infantis mitigates the effect of antibiotics on the microbiota and childhood asthma risk. *Med* **4**, 92–112.e115 (2023).
28. Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M. & Owen, L. J. Dysbiosis of the gut microbiota in disease. *Micro. Ecol. Health Dis.* **26**, 26191–26191 (2015).
29. Stokholm, J. et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat. Commun.* **9**, 141 (2018).
30. Hoskinson, C. et al. Delayed gut microbiota maturation in the first year of life is a hallmark of pediatric allergic disease. *Nat. Commun.* **14**, 4785 (2023).
31. Korpela, K. et al. Cohort profile: Finnish health and early life microbiota (HELMi) longitudinal birth cohort. *Bmj. Open* **9**, 6 (2019).
32. Ebrahim, G. J. WHO child growth standards. Growth velocity based on weight, length and head circumference. Methods and development. *J. Trop. Pediatr.***56**, 136–136 (2009).
33. Laursen, M. F. et al. Having older siblings is associated with gut microbiota development during early childhood. *BMC Microbiol.* **15**, 154 (2015).
34. Kim, H., Sitarik, A. R., Woodcroft, K., Johnson, C. C. & Zoratti, E. Birth mode, breastfeeding, pet exposure, and antibiotic use: associations with the gut microbiome and sensitization in children. *Curr. Allergy Asthma Rep.* **19**, 22 (2019).
35. Nagai, M. et al. High body temperature increases gut microbiota-dependent host resistance to influenza A virus and SARS-CoV-2 infection. *Nat. Commun.* **14**, 3863 (2023).
36. Budden, K. F. et al. Emerging pathogenic links between microbiota and the gut–lung axis. *Nat. Rev. Microbiol.* **15**, 55–63 (2017).
37. Reyman, M. et al. Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. *Nat. Commun.* **10**, 4997 (2019).
38. Schuijt, T. J. et al. The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. *Gut* **65**, 575 (2016).
39. Underwood, M. A., German, J. B., Lebrilla, C. B. & Mills, D. A. Bifidobacterium longum subspecies infantis: champion colonizer of the infant gut. *Pediatr. Res.* **77**, 229–235 (2015).
40. Korpela, K. et al. Probiotic supplementation restores normal microbiota composition and function in antibiotic-treated and in caesarean-born infants. *Microbiome* **6**, 182 (2018).
41. Kuitunen, M. et al. Probiotics prevent IgE-associated allergy until age 5 years in cesarean-delivered children but not in the total cohort. *J. Allergy Clin. Immunol.* **123**, 335–341 (2009).
42. Corrêa, N. B. O., Péret Filho, L. A., Penna, F. J., Lima, F. M. L. S. & Nicoli, J. R. A Randomized formula controlled trial of bifidobacterium lactis and streptococcus thermophilus for prevention of antibiotic-associated diarrhea in infants. *J. Clin. Gastroenterol.* **39**, 385–9 (2005).
43. Saturio, S. et al. Role of Bifidobacteria on Infant Health. *Microorganisms* **9**, 2415 (2021).
44. Kumar, H. et al. The bifidogenic effect revisited—ecology and health perspectives of bifidobacterial colonization in early life. *Microorganisms* **8**, 1855 (2020).
45. Marteau, P., Guyonnet, D., Lafaye de Micheaux, P. & Gelu, S. A randomized, double-blind, controlled study and pooled analysis of two identical trials of fermented milk containing probiotic Bifidobacterium lactis CNCM I-2494 in healthy women reporting minor digestive symptoms. *Neurogastroenterol. Motil.* **25**, 331–e252 (2013).
46. Agrawal, A. et al. Clinical trial: the effects of a fermented milk product containing Bifidobacterium lactis DN-173 010 on abdominal distension and gastrointestinal transit in irritable bowel syndrome with constipation. *Aliment. Pharm. Ther.* **29**, 104–114 (2009).
47. Hidalgo-Cantabrana, C. et al. Bifidobacteria and their health-promoting effects. *Microbiol. Spect.* https://doi.org/10.1128/microbiolspec.bad-0010-2016 (2017).
48. Tamana, S. K. et al. Bacteroides-dominant gut microbiome of late infancy is associated with enhanced neurodevelopment. *Gut Microbes* **13**, 1930875 (2021).
49. Korpela, K. et al. Fucosylated oligosaccharides in mother's milk alleviate the effects of caesarean birth on infant gut microbiota. *Sci. Rep.* **8**, 13757 (2018).
50. Ojima, M. N. et al. Priority effects shape the structure of infant-type Bifidobacterium communities on human milk oligosaccharides. *ISME J.* **16**, 2265–2279 (2022).
51. Dedon, L. R. et al. Fucosylated human milk oligosaccharides drive structure-specific syntrophy between Bifidobacterium infantis and Eubacterium hallii within a modeled infant gut microbiome. *Mol. Nutrition Food Res.* **67**, e2200851 (2023).
52. Moens, F., Verce, M. & De Vuyst, L. Lactate- and acetate-based cross-feeding interactions between selected strains of lactobacilli, bifidobacteria and colon bacteria in the presence of inulin-type fructans. *Int. J. Food Microbiol.* **241**, 225–236 (2017).
53. Cheng, H. Y. et al. Evaluation of stool short chain fatty acids profiles in the first year of life with childhood atopy-related outcomes. *Front. Allergy* **3**, 873168 (2022).
54. Matharu, D. et al. Bacteroides abundance drives birth mode dependent infant gut microbiota developmental trajectories. *Front. Microbiol.* **13**, 953475 (2022).
55. Lazarenko, L. et al. Antagonistic action of lactobacilli and bifidobacteria in relation to staphylococcus aureus and their influence on the immune response in cases of intravaginal staphylococcosis in mice. *Probiotics Antimicrob. Proteins* **4**, 78–89 (2012).
56. Jokela, R. et al. Quantitative insights into effects of intrapartum antibiotics and birth mode on infant gut microbiota in relation to well-being during the first year of life. *Gut Microbes.* **14**, 2095775 (2022).
57. Laforest-Lapointe, I. & Arrieta, M. C. Patterns of early-life gut microbial colonization during human immune development: an ecological perspective. *Front. Immunol.* **8**, 788 (2017).
58. Mitchell, C. M. et al. Delivery mode affects stability of early infant gut microbiota. *Cell Rep. Med.* **1**, 100156 (2020).
59. Maldonado, J. et al. Human milk probiotic lactobacillus fermentum CECT5716 reduces the incidence of gastrointestinal and upper respiratory tract infections in infants. *J. Pediatric Gastroenterol. Nut.* **54**, 55–61 (2012).
60. Pham, V. T., Lacroix, C., Braegger, C. P. & Chassard, C. Early colonization of functional groups of microbes in the infant gut. *Environ. Microbiol.* **18**, 2246–2258 (2016).
61. Gaspar, C. et al. Bacteriocin production of the probiotic Lactobacillus acidophilus KS400. *AMB Express* **8**, 153–153 (2018).

62. dos Reis Buzzo Zermiani, A. P. et al. Evidence of Lactobacillus reuteri to reduce colic in breastfed babies: Systematic review and meta-analysis. *Complement. Ther. Med.* **63**, 102781 (2021).
63. Nilsen, M. et al. Butyrate levels in the transition from an infant- to an adult-like gut microbiota correlate with bacterial networks associated with Eubacterium rectale and ruminococcus Gnavus. *Genes* **11**, 1245 (2020).
64. Belzer, C. et al. Microbial metabolic networks at the mucus layer lead to diet-independent butyrate and vitamin B12 production by intestinal symbionts. *mBio* **8**, :e00770-17 (2017).
65. Gupta, V. K. et al. A predictive index for health status using species-level gut microbiome profiling. *Nat. Commun.* **11**, 4635 (2020).
66. Gacesa, R. et al. Environmental factors shaping the gut microbiome in a Dutch population. *Nature* **604**, 732–739 (2022).
67. Zhu, J. et al. Statistical modeling of gut microbiota for personalized health status monitoring. *Microbiome* **11**, 184 (2023).
68. Delost, M. D. *Introduction to Diagnostic Microbiology* 2nd edn, Vol. 600 (Jones & Bartlett Learning, 1997).
69. Pham, V. T., Lacroix, C., Braegger, C. P. & Chassard, C. Lactate-utilizing community is associated with gut microbiota dysbiosis in colicky infants. *Sci. Rep.* **7**, 11176 (2017).
70. Penders, J. et al. New insights into the hygiene hypothesis in allergic diseases: mediation of sibling and birth mode effects by the gut microbiota. *Gut Microbes* **5**, 239–244 (2014).
71. Jokela, R. et al. Sources of gut microbiota variation in a large longitudinal finnish infant cohort. *eBioMedicine* **94**, 104695 (2023).
72. Strachan, D. P. Hay fever, hygiene, and household size. *Bmj* **299**, 1259–1260 (1989).
73. Gao, Y. et al. Gut microbiota maturity mediates the protective effect of siblings on food allergy. *J. Allergy Clin. Immunol.* **152**, 667–675 (2023).
74. Karav, S., Casaburi, G. & Frese, S. A. Reduced colonic mucin degradation in breastfed infants colonized by Bifidobacterium longum subsp. infantis EVC001. *FEBS Open Bio.* **8**, 1649–1657 (2018).
75. Quin, C. & Gibson, D. L. Human behavior, not race or geography, is the strongest predictor of microbial succession in the gut bacteriome of infants. *Gut Microbes.* **11**, 1143–1171 (2020).
76. Weiland, S. K. et al. Phase II of the international study of asthma and allergies in childhood (ISAAC II): rationale and methods. *Eur. Respir. J.* **24**, 406–412 (2004).
77. Salonen, A. et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J. Microbiol. Methods* **81**, 127–134 (2010).
78. Korpela, K. *mare: Microbiota Analysis in R Easily*. https://github.com/katrikorpela/mare (2016).
79. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
80. Korpela, K. et al. Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nat. Commun.* **7**, 10410 (2016).
81. Cole, J. R. et al. Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**, D633–D642 (2013).
82. Jokela, R. et al. Sources of gut microbiota variation in a large longitudinal Finnish infant cohort. *EBioMedicine* **94**, 104695 (2023).
83. Oksanen, J. et al. *vegan: Community Ecology Package*. https://cran.r-project.org/web/packages/vegan/vegan.pdf (2008).
84. Machler, M., Rousseeuw, P., Struyf, A. & Hubert, M. *Cluster: Cluster Analysis Basics and Extensions*. https://www.researchgate.net/publication/272176869_Cluster_Cluster_Analysis_Basics_and_Extensions (2022).
85. Nagin, D. S. Group-based trajectory modeling: an overview. *Ann. Nutr. Metab.* **65**, 205–210 (2014).
86. Magrini, A. Assessment of agricultural sustainability in European Union countries: a group-based multivariate trajectory approach. *AStA Adv. Stat. Anal.* **106**, 673–703 (2022).
87. *rpart: Recursive Partitioning and Regression Trees*. https://cran.r-project.org/web/packages/rpart/rpart.pdf (2022).
88. Stokholm, J. et al. Cesarean section changes neonatal gut colonization. *J. Allergy Clin. Immunol.* **138**, 881–889.e882 (2016).
89. Lin, S. et al. Fusobacterium nucleatum aggravates ulcerative colitis through promoting gut microbiota dysbiosis and dysmetabolism. *J. Periodontol.* **94**, 405–418 (2023).
90. Hornef, M. Pathogens, commensal symbionts, and pathobionts: discovery and functional effects on the host. *Ilar J.* **56**, 159–162 (2015).
91. Angelakis, E. et al. Treponema species enrich the gut microbiota of traditional rural populations but are absent from urban individuals. *N. Microbes N. Infect.* **27**, 14–21 (2019).
92. Costa, D. & Iraola, G. Pathogenomics of emerging campylobacter species. *Clin. Microbiol. Rev.* **32**, e00072-18. (2019).
93. Moles, L. et al. Serratia marcescens colonization in preterm neonates during their neonatal intensive care unit stay. *Antimicrob. Resist Infect. Control* **8**, 135 (2019).
94. Venables, W. N., Ripley, B. D. & Venables, W. N. *Modern Applied Statistics With S* 4th edn, Vol. 498 (Springer, 2002).
95. nlme. *Linear and Nonlinear Mixed Effects Models*. https://cran.r-project.org/web/packages/nlme/nlme.pdf (2021).
96. Hickman, B. *Wellbeing-Index*. https://github.com/bhick001/Wellbeing-Index (2024).

Acknowledgements

Figure 4a was created with BioRender.com with a full license provided through the University of Helsinki. This study was supported by the doctoral school of Population Health, University of Helsinki, Finland (B.H.). Microbiota sequencing was partly covered by the SIAM Gravitation Grant 024.002.002, of the Netherlands Organization for Scientific Research (NWO). The study was supported by grants from Tekes 329/31/2015 (W.M.d.V. and A.S.), Academy of Finland 1325103 (A.S.), and 339172 (A.P.), Maty and Georg C. Ehrnrooth Foundation (A.S.). Open access funding provided by Helsinki University Library.

Author contributions

W.M.d.V. and A.S. conceptualised the HELMi cohort and together with K.-L.K. and K.K. designed it. K.K., A.S., and W.M.d.V. conceptualized the study. K.K. designed the study. B.H. processed the raw sequence reads. B.H. and K.K. performed the data analysis, data interpretation, figure generation, and wrote the paper. R.J. and A.P. performed metadata curation. All authors read, edited, revised, and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52561-6.

Correspondence and requests for materials should be addressed to Brandon Hickman or Katri Korpela.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.