The composition of the human gut microbiome matures within the first years of life. It has been hypothesized that microbial compositions in this period can cause immune dysregulations and potentially cause asthma. Here we show, by associating gut microbial composition from 16S rRNA gene amplicon sequencing during the first year of life with subsequent risk of asthma in 690 participants, that 1-year-old children with an immature microbial composition have an increased risk of asthma at age 5 years. This association is only apparent among children born to asthmatic mothers, suggesting that lacking microbial stimulation during the first year of life can trigger their inherited asthma risk. Conversely, adequate maturation of the gut microbiome in this period may protect these pre-disposed children.
The first years of our lives represent a period that is critical for susceptibility to environmental exposures. Here, lasting effects may be introduced to the developing immune system through complex host–environment interactions. The human microbiome contains as many as 10^{14} bacteria, similar to the number of cells of an individual. From birth, humans are continuously subjected to multiple exposures that influence microbiome ecology. The composition of the gut microbiome matures within the first years of life, and the microbiome may have the ability to affect host immune maturation. Perturbation of the microbiome during this critical period of development may cause asthma, allergy, and other immunologic disorders. Thus, the microbiome may be an important environmental factor that determines the transition from health to disease.

The neonatal microbial colonization patterns are greatly affected by mode of delivery and use of intrapartum antibiotics, factors which may cause long-term microbial derangements, although the extent of these derangements differs between studies. Birth by cesarean section is a recognized risk factor for asthma, as well as for other immune-mediated diseases in childhood. Furthermore, antibiotic exposure in the first year of life has also been associated with increased asthma risk, again pointing to microbe-mediated mechanisms.

The objective of this study is to analyze the nature of gut colonization patterns during the first year of life, and the associations of these patterns with the later risk of asthma among children from the Copenhagen Prospective Studies on Asthma in Childhood (COPSAC) birth cohort. We show that the gut microbiome, as assessed by β-diversity, relative abundances of genera and microbial maturation at age 1 year, is associated with asthma at age 5 years in the 690 children. This effect is only apparent in children born to asthmatic mothers, and especially characterizes an asthma phenotype also comprising allergic sensitization. This indicates that the development of the gut microbiome during the first year of life can impact the development of childhood asthma, and that adequate microbial maturation in this period may protect pre-disposed children.

**Results**

**Baseline characteristics of the study cohort.** The COPSAC birth cohort has been followed prospectively with deep clinical phenotyping and structured interviews at 11 scheduled visits during the first 5 years of life, with asthma as primary outcome. The characteristics of the entire cohort of 700 children have been described in detail. Of these children, 51% were boys, 57% had at least one older child in the home at birth, 22% were delivered by cesarean section, and 46% were treated with antibiotics during the first year of life. The mean maternal age at delivery was 32.2 years and 26% of the mothers had physician-diagnosed asthma.

**Microbial composition changes during first year of life.** The composition of the gut microbiome changes extensively in early life, with regard to diversity, complexity, and dominant bacterial taxa. Therefore, we examined the compositional changes that occurred in the first year of life, before asthma onset. A total of 1696 fecal samples from 690 children arrived in the lab within 24 h of being produced and were characterized by 16S rRNA gene amplicon sequencing of the V4 region. With a median sequencing depth of 44,827 reads (interquartile range (IQR): 2358–78,208) increasing with age of sample, we identified 3651 unique operational taxonomic units (OTUs), demonstrating a median richness of 116 OTUs per sample, with the dominating genera *Bacteroides, Bifidobacterium,* and *Veillonella.* Alpha-diversity, assessed by the Shannon diversity index, was not different between 1 week (median (IQR), 2.0 (1.5–2.5)) and 1 month (1.9 (1.4–2.5)), but substantially increased at age 1 year (2.8 (2.4–3.3)), (P < 0.001). Similarly, the Chao1 index showed a minor decrease between 1 week (131 (96–175)) and 1 month (121 (78–168)) (P < 0.001) and a substantial increase at age 1 year (295 (203–366)) (P < 0.001). Although significant differences in the population structure (β-diversity), as determined by weighted UniFrac distances, were found between ages 1 week and 1 month (F = 26.5, R^2 = 2.4%, P < 0.001), the greatest change in structure was between ages 1 month and 1 year (F = 558.5, R^2 = 33.2%, P < 0.001, Fig. 1a). To evaluate which bacterial genera were involved in these observed temporal differences in α- and β-diversity, we examined the relative abundances of the most prevalent taxa. Each of the 20 most abundant genera changed significantly in relative abundance during the first year of life (Kruskal–Wallis test, all P-values < 0.01) with the largest differences observed from age 1 month to 1 year (Supplementary Fig. 1).

**Microbial community types and maturation are age-determined.** To further describe the compositional differences in the microbial populations, we applied a clustering method (Partitioning around medoids (PAM) clustering) to separate all the fecal samples by their weighted UniFrac distances taking microbial phylogeny and abundances into account, but without including the time point in the computational model (Fig. 1b). This approach divided the samples into two distinct microbial community types (PAM clusters) (silhouette width, 0.30, largely representing the child’s age at sampling, with PAM cluster 1 (N = 1019) mainly composed of the 1-week and 1-month samples and PAM cluster 2 (N = 677) chiefly composed of the 1-year samples (Fig. 1c). The five most discriminating indicator OTUs for each cluster were identified for PAM cluster 1 as *Enterobacteriaceae*, *Staphylococcus*, *Streptococcus*, *Bifidobacterium* and *Enterococcus,* and for PAM cluster 2 as *Faecalibacterium*, *Bacteroides*, and *Anaerostipes.* Using the Shannon diversity index, we identified higher α-diversity, as well as a microbiome dominated by the phylum Bacteroidetes in PAM cluster 2 compared to PAM cluster 1, which was dominated by Proteobacteria (Supplementary Fig. 2). As such, these PAM clusters likely represent the age-related maturation of the intestinal microbial populations.

To determine whether the PAM cluster transition at age 1 year was dependent on environmental factors, a wide range of exposures (including delivery method and antibiotics in the first year of life) were examined. Among these, only the presence of older children in the home from birth was significantly associated with the clusters; of the 34 children with a PAM cluster 1 composition at age 1 year, only 24% (N = 8) had older children in the home compared to 57% (N = 334) of the 589 children in PAM cluster 2 (χ² test P < 0.001). The microbial populations by means of β-diversity at age 1 month was not associated with PAM cluster at age 1 year.

To further explore the microbial maturation process over time, we calculated microbiota-by-age z-scores (MAZ) for all samples. The model is trained from the microbial composition on a data set with known sample age, and afterward predicts the microbial age for each sample based solely on the microbial composition. Microbiota age increased significantly for each adjacent time point (Wilcoxon test, P-values < 0.001) (Supplementary Fig. 3).

**Microbial population structure associates with later asthma.** Next, we sought to examine whether associations existed between...
the overall microbial composition at the different time points and the later development of asthma. Current asthma at age 5 years was used as the primary end-point, as it represents a strong persistent clinical phenotype; many children experience episodes of asthma-like symptoms in the first years of life, but outgrow their asthma symptoms before school-age\(^\text{35}\). Among the 648 children with follow-up to age 5 years, the prevalence of ongoing asthma at age 5 was 9\% (\(N=60\)). There were no significant associations between \(\alpha\)-diversity (Shannon diversity and Chao1 indices) at any time point and asthma risk, also after adjustment for potential confounders (older siblings, duration of exclusive breastfeeding, hospitalization after birth, antibiotic use, and delivery mode). There were no significant associations between \(\beta\)-diversity at the two earliest time points; however, the microbial populations were significantly different at 1 year in children who had asthma at age 5 compared to non-asthmatics (\(F=3.4, R^2=0.6\%\), \(P=0.003\)). To further explore this finding, we stratified the samples at age 1 year according to maternal asthma status. The microbiome-asthma association was found only in the 147 children born to asthmatic mothers (\(F=6.3, R^2=4.2\%\), \(P<0.001\)) but not in children of non-asthmatic mothers (\(F=0.7, R^2=0.2\%\), \(P=0.65\)), demonstrating a significant interaction (\(P=0.003\)) (Fig. 2). In a test of sensitivity, we adjusted these analyses for potential confounders (older siblings, duration of exclusive breastfeeding, hospitalization after birth, antibiotic use, and delivery mode). The significance levels remained essentially unchanged (All children: \(F=3.5, R^2=0.6\%\), \(P<0.001\); asthmatic mother: \(F=6.2, R^2=4.1\%\), \(P<0.001\); non-asthmatic mother: \(F=0.7, R^2=0.2\%\), \(P=0.647\)). There were no differences in relation to asthma status of the father. Neither maternal nor paternal asthma status was associated with the \(\alpha\)- or \(\beta\)-diversity at any time point (maternal asthma vs. \(\beta\)-diversity: Supplementary Fig. 4). However, that the microbial composition was not affected by maternal asthma status suggests that only susceptible children, exposed to inappropriate microbial stimulation during the first year of life, may express their inherited asthma risk. In a test of sensitivity, we also examined associations between population structure at age 1 year in children born asthmatic mothers and asthma risk using other \(\beta\)-diversity indices that are not biased by \(\beta\)-diversity: Supplementary (Fig. 4). However, the microbial composition was not affected by maternal asthma status suggests that only susceptible children, exposed to inappropriate microbial stimulation during the first year of life, may express their inherited asthma risk. In a test of sensitivity, we also examined associations between population structure at age 1 year in children born asthmatic mothers and asthma risk using other \(\beta\)-diversity indices that are not biased by the dominant taxa (unweighted UniFrac and Jensen–Shannon divergence), which yielded similar results (\(P<0.001\)).

Relative abundance at age 1 year associates with later asthma.

To understand whether the later asthma risk was affected by specific bacterial genera present earlier, we examined the relative abundances of the most common genera. For the 20 most abundant genera present in the 1-year samples, we observed increased risk of asthma at age 5 years associated with higher abundance of *Veillonella* (asthma vs. non-asthma; median relative abundance, 0.94 vs. 0.29\%; Wilcoxon rank-sum test, \(P=0.035\)) and with lower abundance of *Roseburia* (0.27 vs. 0.66\%; \(P=0.001\)).
The majority of these genera were negatively associated with relative abundance in the age 1-year sample of the genera Faecalibacterium (0.59 vs. 3.27%; P = 0.010), Bifidobacterium (0.47 vs. 2.27%; P = 0.006), Roseburia (0.01 vs. 0.76%; P < 0.001), Alistipes (0.01 vs. 0.39%; P = 0.003), Lachnospiraceae incertae sedis (0.05 vs. 0.19%; P = 0.018), Ruminococcus (<0.01 vs. 0.13%; P = 0.004) and Dialister (<0.01 vs. 0.15%; P = 0.007) and positively correlated only with Veillonella (1.41 vs. 0.23%; P = 0.039) (Fig. 3). The genera that were found in lower abundances at age 1 year among children who later became asthmatics have been considered as determinants of a healthy mature gut composition. The majority of these genera were highly correlated at 1 year, while being negatively correlated with the genus Veillonella (Supplementary Fig. 5). None of the 20 most abundant genera at 1 week or 1 month were associated with asthma development. To further examine the associations between the microbial composition at age 1 year and asthma at age 5 in children born to asthmatic mothers, a cross-validated sparse PLS model was constructed to identify jointly contributing taxa at age 1 year that would predict later asthma in these children. This resulted in a one-component model based on the relative abundances of 60 genera (Supplementary Fig. 6). The model demonstrated a high predictive capacity for asthma (cross-validated AUC 0.76) and the many contributing taxa suggest a global delayed microbial maturation at age 1 year in children with asthma at age 5.

Community types at age 1 year associates with later asthma. To formally test whether the maturation of the microbiome was a determinant for asthma development, we used our microbial community types (PAM clusters) as indicators of microbial maturation. After excluding children without full 5-year follow-up, transient asthma and diagnosis before the 1-year sample (N = 105), the risk of developing persistent asthma if the child’s microbiome remained in PAM cluster 1 at 1 year of age (N = 28) was compared to children with transition to PAM cluster 2 (N = 490), and all analyses were adjusted for the presence of older children (Fig. 4). During the first 5 years of life, the risk of developing persistent asthma was increased (adjusted hazard ratio aHR 2.87 (1.25–6.55), P = 0.013) if the microbiome remained in PAM cluster 1 at age 1. This effect was driven solely by the 120 children born to asthmatic mothers (PAM1, N = 7 vs. PAM2, N = 113) (aHR 12.99 (4.17–40.51), P < 0.001), whereas there was no microbial effect on asthma development for the 398 children of non-asthmatic mothers (PAM1, N = 21 vs. PAM2, N = 377) (aHR 0.56 (0.07–4.16), P = 0.57), demonstrating significant interaction, P = 0.011. In a test of sensitivity, we adjusted these analyses further by including other potential confounders (older siblings, duration of exclusive breastfeeding, hospitalization after birth, antibiotic use, and delivery mode). All estimates and significance levels remained essentially unchanged (all children: aHR 2.91 (1.25–6.79), P = 0.013; asthmatic mother: aHR 10.92 (3.44–34.67), P < 0.001; non-asthmatic mother: aHR 0.56 (0.07–4.23), P = 0.572). We found no associations of PAM clusters with the transient asthma phenotype. These analyses provide further evidence that having the more immature gut microbial composition at age 1 year may be a trigger for asthma development in susceptible children.

Low 1-year microbial maturity associates with later asthma. We then applied another validated method for determining the early life maturation of the intestinal microbial populations using the MAZ score. The microbial maturity (predicted microbiota age median microbiota age) and MAZ (microbial maturity/SD) were used as metrics. We sought to determine whether a lower MAZ score at age 1 year was associated with later asthma. During the first 5 years of life, the risk of developing persistent asthma was increased with low maturity (below median) in MAZ at 1 year (low maturity, N = 257 vs. high maturity, N = 261) (HR 1.77 (1.02–3.07), P = 0.043). Low microbial maturity was only associated with later asthma in children born to asthmatic mothers (low maturity, N = 63 vs. high maturity, N = 57) (HR 6.53 (1.93–22.06), P = 0.003), and not in children of non-asthmatic mothers (low maturity, N = 194 vs. high maturity, N = 204) (HR 0.92 (0.46–1.84), P = 0.81), demonstrating significant
interaction, \( P = 0.006 \) (Supplementary Fig. 7). In total, adequate maturation of the gut microbiome in this period appears critical for healthy (non-asthmatic) development and may protect children with pre-dispositions.

Low microbial maturity associates with number of asthma-like episodes. We then examined a secondary asthmatic end-point, the total number of asthma-like episodes in the first 3 years of life, which is a marker of disease burden. Children with microbiome

Fig. 3 Relative abundances in the 1-year fecal samples associate with later asthma. Comparison among the 20 most abundant bacterial genera. Relative abundance of each genus is shown with respect to asthma at age 5 years in all children, and stratified by maternal asthma. \( P \)-values correspond to Wilcoxon rank-sum tests of the relative abundances, with significant values (\( P < 0.05 \)) bolded. FDR limits were calculated for the comparisons: Bonferroni (all: \( P < 0.0025 \)), Benjamini & Hochberg (all children: \( P < 0.0024 \), asthmatic mother: \( P < 0.0102 \), non-asthmatic mother: \( P < 0.05 \)). A pseudocount (+1e -06) was added to all abundances for the log-scale presentation. The black dots indicate median values and the abundances are colored according to the asthmatic (orange) (\( N = 58 \)) or non-asthmatic (green) (\( N = 531 \)) status of the child at age 5 years.

Fig. 4 Kaplan–Meier plots illustrate risk of asthma in the first 5 years of life according to Partitioning Around Medoids (PAM) cluster at 1 year. Comparisons shown of all children and stratified by maternal asthma, demonstrating interaction between PAM cluster and maternal asthma (\( P = 0.01 \)). Associations are quantified by Cox proportional hazards regression, adjusted for presence of older children in the home (aHR, adjusted Hazard Ratio). Children with a transient asthmatic phenotype (remission before age 5 years) are excluded. Curves are colored according to cluster; All: PAM cluster 1 (red) (\( N = 28 \)), PAM cluster 2 (blue) (\( N = 490 \)). Subsets include: asthmatic mother, PAM cluster 1 (red) (\( N = 7 \)), PAM cluster 2 (blue) (\( N = 113 \)); non-asthmatic mother: PAM cluster 1 (red) (\( N = 21 \)), PAM cluster 2 (blue) (\( N = 377 \)).
Figure 5: β-diversity at 1 year associates with asthma and allergic sensitization. PCoA plots of weighted UniFrac distances. Microbial compositions are assessed in relation to a child’s asthma status at age 5 years and allergic sensitization before age 18 months, and stratified by maternal asthma. P-values correspond to Adonis PERMANOVA tests. Ellipses demonstrate the mean ± 1 SD of children, who at age 5 years were: asthmatic and sensitized (red) (N = 439), non-asthmatic and sensitized (blue) (N = 45), or non-asthmatic and not sensitized (green) (N = 465). Subsets include: asthmatic mother, asthmatic and sensitized (red) (N = 59), non-asthmatic and not sensitized (purple) (N = 19), non-asthmatic and sensitized (blue) (N = 15), and non-asthmatic and not sensitized (green) (N = 103); non-asthmatic mother: asthmatic and sensitized (red) (N = 7), asthmatic and not sensitized (purple) (N = 26), non-asthmatic and sensitized (blue) (N = 44), and non-asthmatic and not sensitized (green) (N = 362).

The associations remain robust at higher sequencing depth. To test the robustness of these findings, the analyses were stratified by sequencing depth to only include samples with more than 10,000 reads. The PAM clustering, MAZ, and association results with asthma and episodes of troublesome lung symptoms all were robust with reference to this more stringent inclusion criterion. Furthermore, we found no significant differences in sequencing depth at 1 year between children, who did or did not develop asthma (median (IQR)); asthma vs. non-asthma, 54,260 (12,018–74,971) vs. 51,940 (12,010–73,050) (P = 0.8); adjustment of the results for sequencing depth resulted in essentially no difference.

Microbial populations describe asthma with sensitization. To explore whether the results were further subdivided by allergic sensitization in childhood, the microbial populations by means of β-diversity at age 1 year were significantly different whether the child had asthma or not with or without sensitization (four categories of comparison) (P = 1.7, R^2 = 0.9% (P = 0.020).

Again, this was mainly apparent among children born to asthmatic mothers (F = 2.9, R^2 = 5.9%, P < 0.001) with no significant effects among children born to non-asthmatic mothers F = 0.7, R^2 = 0.5%, (P = 0.887). Especially, the phenotype of asthma with allergic sensitization was associated with skewed β-diversity (Fig. 5).

Discussion
With the primary objective of making accurate asthma diagnoses and comprehensive phenotyping, we found that the gut microbiome at age 1 year was associated with asthma at age 5 in a prospective cohort of 700 children. Parallel observations were apparent for overall β-diversity, relative abundances of genera, and maturation by means of PAM clustering and MAZ. The effect was only apparent in children born to asthmatic mothers, but the composition and maturation of the gut microbiome in early life was not affected by maternal asthma status. Because of the observational design of the study, we cannot unravel causality or directionality, but our findings indicate that children born to asthmatic mothers harbor a susceptibility to influences from the gut microbiome and that adequate maturation in this period may protect these pre-disposed children.

The main study strength is the design of the COPSAC2010 birth cohort, which is a replication and extension of the COPSAC2000 cohort with regular longitudinal clinical follow-up, meticulous daily symptom recording, and acute care visits to the clinical research site, which has a long record of performing clinical follow-up in cohort studies. All diagnoses and treatments of asthma were performed solely by COPSAC physicians. This ensured highly specific end-points and accurate assessment of time of disease onset. Importantly, this allowed for discrimination between persistent and transient asthma phenotypes. Asthma at age 5 years is a robust phenotype compared with wheezing in the first few years of life, which for most children merely describes a transient phenotype, especially when accompanied by increased early life severity and sensitization. Collecting fecal samples longitudinally from age 1 week allowed for description of the development of the microbial populations during the first year of life.

Our study is by design observational, which limits assessment of causal relationships. However, the number of participants allows for stratification of the data by physician-diagnosed
maternal asthma, which suggests potential mechanisms. Our observation also provides an initial indication of which children might benefit from microbial manipulation in the first year of life. Although the 16S rRNA gene amplicon sequencing technique provides a thorough description of the overall microbial populations, with sequencing of only the V4 region, the method only allows for robust classification of the individual taxa at the genus level. This limits our interpretation of the results, as we cannot describe the direct effect of specific species or strains. Furthermore, we cannot rule out that other gut microbiome components, which have not been characterized in the study (e.g., gut viral and/or fungal colonization), could be partly or fully responsible for some of the observed effects.

Our results suggest that the development of persistent asthma in childhood may be mediated through a skewed maturation of the gut microbiome in the first year of life, predominantly in children born to asthmatic mothers. Since the hygiene hypothesis was first described, multiple studies have linked indirect putative causes of microbial alterations in the first years of life with lower risk of immune-related diseases. Laboratory models have also provided proof of the microbiome’s ability to modulate immune functions of the host. Delivery mode has been studied intensively and the associations of cesarean section-birth with asthma and other chronic childhood disorders have been largely consistent. Furthermore, delivery by cesarean section greatly influences the earliest microbiome and its development. Antibiotics in the first year of life have also been confirmed both to perturb the gut microbiome and its development. A study from the Canadian Healthy Infant Longitudinal Development (CHILD) cohort identified associations between lower Faecalibacterium and Lachnospiraceae (including Lachnospiraceae incertae sedis and Roseburia) and Ruminococcaceae (including Ruminococcus and Faecalibacterium) in the first year of life. The lower relative abundances of these genera in children who later become asthmatic could indicate overall delayed microbial maturation. A study from the Canadian Healthy Infant Longitudinal Development (CHILD) cohort identified associations between lower Faecalibacterium and Lachnospiraceae at age 3 months and an early allergic wheezy phenotype at age 1 year, consistent with our findings. However, opposite to our study, they identified Veillonella as being protective. This dichotomy could be explained by differences in the disease phenotypes studied, since early life wheeze is a heterogeneous group not directly comparable with the 5-year asthma phenotype we studied. However, the protective association of Veillonella in the CHILD study was not apparent in the 1-year samples, which is similar to our findings. In the genus Faecalibacterium, the only known species is F. prausnitzii, which is a butyrate-producing bacterium, considered to have anti-inflammatory activities. Conversely, lack of F. prausnitzii is a marker for inflammatory bowel disease, obesity, and diabetes. We found Faecalibacterium as the most specific indicator for the more mature PAM cluster 2 and in lower abundance among children who later became asthmatic. This, and the association with the age-related microbiota score (MAZ), are consistent with a potential causal role, and could suggest a mechanism of delayed maturity influencing immune development in susceptible children. For decades, Bifidobacterium has been recognized as a potential protector against allergic disease and lack of Ruminococcus in earliest life has also been associated with allergic sensitization. The lack of association between bacterial taxa at 1 week and 1 month and later asthma points toward microbial alterations going on between 1 month and 1 year. A deviation might have been observed already at 3 months as suggested in the CHILD study. Furthermore, a skewed microbial composition at age 1 year was mainly associated with an allergic asthmatic phenotype, which is also in line with findings from the CHILD study. The univariate genus-level relative abundance results were further expanded in a multivariate supervised model, again pointing toward differences in the entire bacterial community structure. For children born to asthmatic mothers, we demonstrated a cross-validated AUC of 0.76 for predicting asthma at age 5 years from the microbial composition at 1 year. In comparison, a large study using genetic risk scores of the highest-ranked asthma risk alleles demonstrated an AUC of 0.54–0.57 for prediction of school-age asthma, and an AUC of 0.59 for preschool persistent wheeze.

Maternal asthma status did not affect the microbial populations of the children and therefore did not confound our results. Maternal asthma was however, a key effect modifier between the microbiome and asthma risk in our study, which points to susceptibility to host–microbial interactions specifically for these children. Such susceptibility could arise from an inborn immune deviation determined by maternal asthma status, as we have reported in this cohort.

Our results suggest that the lack of maturation of the gut microbiome in the first year of life is the critical determinant for increased asthma risk as illustrated by the microbial community type analyses. Having older siblings was the only significant determinant for belonging in the more mature PAM cluster 2 at age 1 year, and is consistent with a microbial mechanism for the original hygiene hypothesis. We speculate that microbial transfer from the older child may advance maturation of the gut composition and, through appropriate stimulation of the developing immune system, cause protection from asthma in otherwise susceptible children.

Probiotic intervention trials in pregnancy and in early life have so far failed to prove effective in reducing asthma risk. However, this could be due to species selection or inadequate (single or few species) compositions of the probiotics. It can be speculated that the global microbial population and delayed maturation in the first year of life matters more than individual taxa. Our results suggest possible beneficial effects of at least seven of the most abundant genera among the children born to asthmatic mothers. Because our results are observational and most of the taxa associated with lower asthma risk demonstrate collinearity, we are not able to pinpoint specific causal taxa; however, the multivariate model illustrates many jointly contributing taxa and not a single risk taxon, consistent with a global effect of the microbiota. Therefore, contributing bacterial strains must be identified, potential mechanisms validated, and immunological features determined in future experimental models. We speculate that a mixture of the strains to be identified could form the basis for a next-generation probiotic that could be tested in a randomized controlled trial to children at risk.

Our results suggest that the maturation of the gut microbiome during the first year of life has an important role in the development of childhood asthma, especially in children born to asthmatic mothers, and that older children in the home may help advance the maturation. Our results suggest potential beneficial effects of specific microbial supplementation in the first year of life for children at high risk for developing asthma.
Methods

Ethics statement. The study was conducted in accordance with the guiding principles of the Declaration of Helsinki and was approved by The National Committee on Health Research Ethics (H-2008-093) and the Danish Data Protection Agency (2015-41-3696). Both parents gave written informed consent before enrollment.

Study population. The COPSAC2010 cohort is a population-based birth cohort of 706 children recruited in pregnancy and has been followed prospectively at the COPSAC research unit with daily clinical phenotyping at 11 scheduled visits during the first 5 years of life. Of these 690 had a feal sample characterized. The study pediatricians collected all information during these clinical visits scheduled at 1 week, 1, 3, 6, 12, 18, 24, 30, and 36 months, and yearly thereafter. Additional acute care visits were arranged whenever the children experienced lung or skin symptoms. The symptom burden during visits was captured with daily diary cards monitoring: significant troublesome lung symptoms including components of cough, wheeze, and dyspnea; skin symptoms; and respiratory infections. The study pediatricians were the only physicians responsible for diagnosis and treatment of asthma, allergy, and eczema adhering to predefined algorithms.

Primary end-point. Asthma was diagnosed based on a previously detailed quantitative symptom algorithm requiring all of the following criteria: (1) verified diary records of five episodes of troublesome lung symptoms within 6 months, each lasting at least 3 consecutive days; (2) symptoms typical of asthma including exercise-induced symptoms, prolonged nocturnal cough, and/or persistent cough outside the immediate environment; (3) need for intermittent inhaled β2-agonist; and (4) response to a 3-month course of inhaled corticosteroids and relapse upon ending treatment. Remission was defined by 12 months without relapse upon cessation of inhaled corticosteroid treatment. For analyses, we used the cross-sectional ongoing asthma diagnosis at age 5 years and the time to disease relapse upon cessation of inhaled corticosteroid treatment. Data processing: Fastq-files demultiplexed by the MiSeq Controller Software (Illumina Inc.) were trimmed for amplification primers, diversity spaces, and sequencing adapters using biopieces (www.biopieces.org), mate-paired and quality filtered (USEARCH v7.0.1090)58, UPARSE59,60 was used for OTU clustering as recommended, in particular removing singletones after dereplication. Chimera checking was performed with UCHIME61 against Mothur supplied version of the RDP9 PDS database. Repetitive sequences were removed and sequence classification was performed using the Mothur v1.25.0 wang function62, FastTree63 in nucleotide mode and Mothur align.seq function64 were used to construct a phylogenetic tree. Alignments were built with reference to the 2013 version of Greengenes65.

Statistics and data analysis. All data analysis was performed in the statistical software package R version 3.3.0, with the package phyloseq66 to handle the microbiome data. The Shannon diversity index and the Chao1 index were used as measures of diversity within individual gut microbiota (α-diversity). Individu- al diversity metrics (β-diversity) were computed as weighted UniFrac67 distances after adding a pseudocount (+1) to all OTU counts and log-transforming the OTU tables. For sensitivity analyses unweighted UniFrac and Jensen–Shannon divergence were used as other metrics of β-diversity. Differences in α-diversity between groups were compared using Wilcoxon rank-sum test. Linear regression was used to visualize principal coordinates analysis (PCoA) plots, in which the PCoA ordinations were calculated based on all the samples visualized in the specific plot and subsequently stratified. β-diversity was tested for inference using permutation multivariate analysis of variance (PERMANOVA) (PERMANOVA+ (Adonis from the vegan package with 999 permutations))68. In the interaction analyses, we included an interaction term (e.g. asthma * asthma_status_mother). Differences in relative abundances at the genus level were analyzed using Wilcoxon rank-sum (2 levels) or Kruskal–Wallis Test (>2 levels). We used sparse partial least squares (PLS) modeling of asthma at age 5 years after filtering genera (prevalence >5% of chil- dren) or mean relative abundance (MRAW) and log-transforming abundances, using half the lowest nonzero value as a pseudocount. We selected the optimal number of input variables using repeated 10-fold cross-validation of the area under the curve (AUC) statistic to avoid overfitting. The final model was chosen by the highest median AUC value. The predicted values of left out folds were combined to a PLS score. Loading of each genus in the model was in the figure drawn with standard deviations across repeats as error bars, ranked by relative abundance and colored by phylum. Clustering analysis was performed using partitioning around medoids (PAM) clustering from the package cluster69. The Silhouette index was calculated from weighted UniFrac distances was used for determining the optimal number of clusters in the data. Indicator OTUs were identified using the function multipatt from the packageldspecies70. Microbiota-by-age z-scores (MAZ) were calculated with non-asthmatics serving as the training set, with calculation of: microbial maturity (MM) = |predicted microbiota age − median microbiota age; and MAZ = (MM/SD) of predicted microbiota age as metrics. Associations between principal PCs and MAZ and asthma were assessed by Kaplan-Meier analysis and quantified by Cox proportional hazards regression (P-values correspond to Wald tests). The children were retained in the analysis from birth until age of asthma.

Data processing: Fastq-files demultiplexed by the MiSeq Controller Software (Illumina Inc.) were trimmed for amplification primers, diversity spaces, and sequencing adapters using biopieces (www.biopieces.org), mate-paired and quality filtered (USEARCH v7.0.1090)58, UPARSE59,60 was used for OTU clustering as recommended, in particular removing singletones after dereplication. Chimera checking was performed with UCHIME61 against Mothur supplied version of the RDP9 PDS database. Repetitive sequences were removed and sequence classification was performed using the Mothur v1.25.0 wang function62, FastTree63 in nucleotide mode and Mothur align.seq function64 were used to construct a phylogenetic tree. Alignments were built with reference to the 2013 version of Greengenes65.

Statistics and data analysis. All data analysis was performed in the statistical software package R version 3.3.0, with the package phyloseq66 to handle the microbiome data. The Shannon diversity index and the Chao1 index were used as measures of diversity within individual gut microbiota (α-diversity). Individual diversity metrics (β-diversity) were computed as weighted UniFrac67 distances after adding a pseudocount (+1) to all OTU counts and log-transforming the OTU tables. For sensitivity analyses unweighted UniFrac and Jensen–Shannon divergence were used as other metrics of β-diversity. Differences in α-diversity between groups were compared using Wilcoxon rank-sum test. Linear regression was used to visualize principal coordinates analysis (PCoA) plots, in which the PCoA ordinations were calculated based on all the samples visualized in the specific plot and subsequently stratified. β-diversity was tested for inference using permutation multivariate analysis of variance (PERMANOVA) (Adonis from the vegan package with 999 permutations))68. In the interaction analyses, we included an interaction term (e.g. asthma * asthma_status_mother). Differences in relative abundances at the genus level were analyzed using Wilcoxon rank-sum (2 levels) or Kruskal–Wallis Test (>2 levels). We used sparse partial least squares (PLS) modeling of asthma at age 5 years after filtering genera (prevalence >5% of children) or mean relative abundance (MRAW) and log-transforming abundances, using half the lowest nonzero value as a pseudocount. We selected the optimal number of input variables using repeated 10-fold cross-validation of the area under the curve (AUC) statistic to avoid overfitting. The final model was chosen by the highest median AUC value. The predicted values of left out folds were combined to a PLS score. Loading of each genus in the model was in the figure drawn with standard deviations across repeats as error bars, ranked by relative abundance and colored by phylum. Clustering analysis was performed using partitioning around medoids (PAM) clustering from the package cluster69. The Silhouette index was calculated from weighted UniFrac distances was used for determining the optimal number of clusters in the data. Indicator OTUs were identified using the function multipatt from the packageldspecies70. Microbiota-by-age z-scores (MAZ) were calculated with non-asthmatics serving as the training set, with calculation of: microbial maturity (MM) = |predicted microbiota age − median microbiota age; and MAZ = (MM/SD) of predicted microbiota age as metrics. Associations between principal PCs and MAZ and asthma were assessed by Kaplan-Meier analysis and quantified by Cox proportional hazards regression (P-values correspond to Wald tests). The children were retained in the analysis from birth until age of asthma.

USA, 0.5 μm primer 515 F, 0.5 μm primer 806 R, and 2 μl template DNA, giving a total volume of 20 μl per sample. Reaction was run in a 2720 thermal cycler (Applied Biosystems®; Life Technologies, CA, USA) thermocycling program: 2 min of denaturation at 94 °C, followed by 30 cycles of 20 s at 94 °C (denaturing), 30 s at 56 °C (annealing), and 40 s at 68 °C (elongation), with a final extension at 68 °C for 5 min. For each plate, a negative template-free control and a positive control containing 2 μl DNA from a known bacterial monoc-
References

1. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13, 260–270 (2012).
2. Prescott, S. L. Early-life environmental determinants of allergic diseases and the wider pandemic of inflammatory noncommunicable diseases. J. Allergy Clin. Immunol. 131, 23–30 (2013).
3. Bisgaard, H., Bennellykke, K. & Stokholm, J. Immune-mediated diseases and microbial exposure in early life. Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol. 44, 475–481 (2014).
4. Eckburg, P. B. Diversity of the human intestinal microbial flora. Science 308, 1691–1695 (2005).
5. Bäckström, T., Fjällberg, K. & Blann, A. How colonization by microorganisms in early life shapes the immune system. Science 352, 539–544 (2016).
6. Vatanen, T. et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. Cell 165, 842–853 (2016).
7. Blaser, M. J. Antibiotic use and its consequences for the normal microbiome. Science 352, 544–545 (2016).
8. Cox, L. M. et al. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell 158, 705–721 (2014).
9. Bisgaard, H. et al. Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. J. Allergy Clin. Immunol. 128, 646–652.e5 (2011).
10. Björksten, B., Sepp, E., Julge, K., Voor, T. & Mäiksaa, M. Allergy development and the intestinal microbiota during the first year of life. J. Allergy Clin. Immunol. 108, 516–520 (2001).
11. Penders, J. et al. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. Gut 56, 661–667 (2007).
12. Egge, M. J. et al. Exposure to environmental microorganisms and childhood asthma. N. Engl. J. Med. 364, 701–709 (2011).
13. Penders, J. et al. New insights into the hygiene hypothesis in allergic diseases: mediation of childhood and birth mode effects by the gut microbiota. Gut Microbes 5, 239–244 (2014).
14. Arrieta, M.-C. et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci. Transl. Med. 7, 307ra152 (2015).
15. Renz, H. et al. Gene-environment interaction in chronic disease: a European Science Foundation Forward Look. J. Allergy Clin. Immunol. 128, 527–549 (2011).
16. Dominguez-Bello, M. G. et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl Acad. Sci. USA 107, 11971–11975 (2010).
17. Stokholm, J. et al. Cesarean section changes neonatal gut colonization. J. Allergy Clin. Immunol. 138, 881–889 (2016).
18. Azad, M. et al. Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: a prospective cohort study. PLoS One 7, e38395 (2012).
19. Chiu, D. M. et al. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. Nat. Med. 23, 314–326 (2017).
20. Sevelsted, A., Stokholm, J. & Bisgaard, H. Risk of asthma from caesarean section depends on membrane rupture. J. Paediatr. 171, e1–e4 (2016).
21. Thavagnanam, S., Fleming, J., Bromley, A., Shields, M. D. & Cardwell, C. R. A meta-analysis of the association between caesarean section and childhood asthma. Clin. Exp. Allergy 38, 629–633 (2008).
22. Sevelsted, A., Stokholm, J., Bønnelykke, K. & Bisgaard, H. Cesarean section and chronic immune disorders. Pediatr. Res. 135, e92–e98 (2015).
23. McKeever, T. M., Smith, C. & Hubbard, R. The Importance of prenatal exposures on the development of allergic disease: a birth cohort study using the west midlands general practice database. Am. J. Respir. Crit. Care Med. 186, 827–832 (2002).
24. Fujimura, K. E. et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. Nat. Med. 22, 1187–1191 (2016).
25. Bisgaard, H. et al. Deep phenotyping of the unselected COPSAC2010 birth cohort study. Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol. 43, 1384–1394 (2013).
26. Arumugam, M. et al. Enterotypes of the human gut microbiome. Nature 473, 185–191 (2011).
27. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. Cluster: Finding clusters in data. (2015).
28. Subramanian, S. et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature 510, 417–421 (2014).
29. Bokulich, N. A. et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci. Transl. Med. 8, 343ra82–343ra82 (2016).
30. Yassour, M. et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci. Transl. Med. 8, 343ra18–343ra18 (2016).
31. Glasses, T., Iyer, S. S., Kasper, D. L. & Blumberg, R. S. How colonization by microbiota in early life shapes the immune system. Science 352, 539–544 (2016).
32. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. Cluster: Finding clusters in data. (2015).
33. Cox, L. M. et al. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell 158, 705–721 (2014).
34. Bisgaard, H. et al. Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. J. Allergy Clin. Immunol. 128, 646–652.e5 (2011).
35. Björksten, B., Sepp, E., Julge, K., Voor, T. & Mäiksaa, M. Allergy development and the intestinal microbiota during the first year of life. J. Allergy Clin. Immunol. 108, 516–520 (2001).
36. Penders, J. et al. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. Gut 56, 661–667 (2007).
37. Egge, M. J. et al. Exposure to environmental microorganisms and childhood asthma. N. Engl. J. Med. 364, 701–709 (2011).
38. Penders, J. et al. New insights into the hygiene hypothesis in allergic diseases: mediation of childhood and birth mode effects by the gut microbiota. Gut Microbes 5, 239–244 (2014).
39. Arrieta, M.-C. et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci. Transl. Med. 7, 307ra152 (2015).
40. Renz, H. et al. Gene-environment interaction in chronic disease: a European Science Foundation Forward Look. J. Allergy Clin. Immunol. 128, 527–549 (2011).
41. Dominguez-Bello, M. G. et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl Acad. Sci. USA 107, 11971–11975 (2010).
42. Stokholm, J. et al. Cesarean section changes neonatal gut colonization. J. Allergy Clin. Immunol. 138, 881–889 (2016).
52. Bisgaard, H. et al. Chromosome 17q21 gene variants are associated with asthma and exacerbations but not atopy in early childhood. Am. J. Respir. Crit. Care Med. 179, 179–185 (2009).

53. Bisgaard, H. et al. Childhood asthma after bacterial colonization of the airway in neonates. N. Engl. J. Med. 357, 1487–1495 (2007).

54. Stokholm, J. et al. Prevalence and predictors of antibiotic administration during pregnancy and birth. PLoS ONE 8, e29323 (2013).

55. Neefs, J. M. & De Wachter, R. A proposal for the secondary structure of a variable area of eukaryotic small ribosomal subunit RNA involving the existence of a pseudoknot. Nucleic Acids Res. 18, 5695–5704 (1990).

56. Sundberg, C. et al. 454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters. FEMS Microbiol. Ecol. 85, 612–626 (2013).

57. Yu, Y., Lee, C., Kim, J. & Hwang, S. Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. Biotechnol. Bioeng. 89, 670–679 (2005).

58. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. Bioinforms. Osf. Engl. 26, 2460–2461 (2010).

59. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583 (2016).

60. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996–998 (2013).

61. Haas, B. J. et al. Chimera 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21, 494–504 (2011).

62. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541 (2009).

63. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with pro.

64. McArdle, B. H. & Anderson, M. J. Fitting multivariate models to community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 73, 1576–1585 (2007).

65. Caporaso, J. G. et al. QIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336 (2010).

66. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for bacterial and archaeal richness in 21 full-scale biogas digesters. FEMS Microbiol. Ecol. 85, 612–626 (2013).

67. De Cáceres, M. & Legendre, P. Associations between species and groups of sites: indices and statistical inference. Ecology 90, 3566–3574 (2009).

Acknowledgements
We express our deepest gratitude to the children and families of the COPSAC2010 cohort study for all their support and commitment. We acknowledge and appreciate the unique efforts of the COPSAC research team. J.S. has received funding from the Alfred Benzon Foundation, the Thrasher Research Fund and the Danish Council for Independent Research. COPSAC is funded by private and public research funds all listed on www. copsac.com. The Lundbeck Foundation; The Danish Ministry of Health; Danish Council for Strategic Research and The Capital Region Research Foundation have provided core support for COPSAC. M.B. was supported in part by R01 DK090989 from the National Institutes of Health and by the Ziffl Family and C&K funds. All authors have agreed that the accuracy and integrity of any part of the work has been appropriately investigated and resolved and all have approved the final version of the manuscript. The corresponding author had full access to the data and had final responsibility for the decision to submit for publication.

Author contributions
The guarantor of the study is H.B., from conception and design to conduct of the study and acquisition of data, data analysis, and interpretation of data. All co-authors have contributed substantially to the analyses and interpretation of the data, and have provided important intellectual input. J.S. was responsible for acquisition, analysis, and interpretation of data and has written the first draft of the manuscript. J.T. contributed with analysis and interpretation. A.D.B. wrote the bioinformatics pipeline.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-02573-2.

Competing interests: MJB discloses his participation on the Scientific Advisory Board of Commence, Inc. The other authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.