Transmission of gut bacteria and viruses from mothers to infants

Sanzhima Garmaeva  
Department of Genetics, University of Groningen and University Medical Center Groningen  
https://orcid.org/0000-0002-0429-833X

Trishla Sinha  
Department of Genetics, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands  
https://orcid.org/0000-0002-0992-7983

Anastasia Gulyaeva  
University Medical Center Groningen

Ranko Gacesa  
University of Groningen and University Medical Center Groningen  
https://orcid.org/0000-0003-2119-0539

Sergio Andreu-Sanchez  
University Medical Center Groningen  
https://orcid.org/0000-0002-3503-9971

Amau Vich Vila  
University of Groningen and University Medical Center Groningen  
https://orcid.org/0000-0003-4691-5583

Lianmin Chen  
The First Affiliated Hospital of Nanjing Medical University

Johanne Spreckels  
University Medical Center Groningen  
https://orcid.org/0000-0002-8711-1736

Siobhan Brushett  
University of Groningen and University Medical Center Groningen

Marloes Kruk  
Department of Genetics, University of Groningen and University Medical Center Groningen.

Jackie Dekens  
Department of Genetics, University of Groningen and University Medical Center Groningen.

Jan Sikkema  
University of Groningen and University Medical Center Groningen

Folkert Kuipers  
University Medical Center Groningen  
https://orcid.org/0000-0003-2518-737X

Andrey Shkoporov  
University College Cork  
https://orcid.org/0000-0002-5547-8672

Colin Hill
Article

Keywords: infant gut virome, mother-to-infant transmission, metagenomes

Posted Date: November 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1079760/v1

License: ☺️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Transmission of gut bacteria and viruses from mothers to infants

Sanzhima Garmaeva1*, Trishla Sinha1*, Anastasia Gulyaeva1, Ranko Gacesa1,2, Sergio Andreu-Sánchez1,3, Arnau Vich Vila1,2, Lianmin Chen1, Johanne E Spreckels1, Siobhan Brushett1,4, Marloes Kruk1, Lifelines NEXT cohort study5, Jackie Dekens1,5, Jan Sikkema5, Folkert Kuipers3, Andrey Shkoporov6,7, Colin Hill6,7, Sicco Scherjon8, Alexander Kurilshikov1, Cisca Wijmenga1, Jingyuan Fu1,3 and Alexandra Zhernakova1#

1Department of Genetics, University of Groningen and University Medical Center Groningen, Groningen, the Netherlands
2Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, Groningen, the Netherlands
3Department of Pediatrics, University of Groningen and University Medical Center Groningen, Groningen, the Netherlands
4Department of Health Sciences, University of Groningen and University Medical Center Groningen, Groningen, the Netherlands
5University Medical Center Groningen, Center for Development and Innovation
6APC Microbiome Ireland, University College Cork, Cork, Ireland
7School of Microbiology, University College Cork, Cork, Ireland
8Department of Obstetrics and Gynecology, University of Groningen and University Medical Center Groningen, Groningen, the Netherlands

*shared first authors: SG and TS

#corresponding author: Alexandra Zhernakova
Abstract

Seeding and development of the gut ecosystem are crucial for health, both in childhood and later in life. While the composition of infant gut bacterial communities has been described, the composition and origin of the infant gut virome remains under-studied. Here, we explore mother-to-infant transmission of bacteria and viruses in 30 mother–infant pairs in a longitudinal collection of faecal samples taken during pregnancy and the first 3 months after birth. We demonstrate that infant bacterial strains resemble maternal strains more than those of unrelated mothers. We quantify viromes using a complementary approach examining both total metagenomes and viral metagenomes. The virome composition is highly consistent between viral and total metagenomes. The infant gut viromes are dominated by active temperate bacteriophages, which are more abundant in infants than mothers (p-value=7.2e-06). We observe that the proportion of shared viruses between maternal and infant gut is only 11.3% when considering the active virome fraction alone, but increases to 37.6% when taking into account temperate phages in the form of prophages. These findings indicate that viruses are vertically transmitted from mothers to infants early in life and that pioneering phages can reach the infant gut via vertical transmission of their bacterial hosts.

Introduction

The human early life gut ecosystem has garnered much interest in recent years because of its links to health and disease later in life, but the core aspects of its origin and development remain less understood. Studies have characterised the development of the infant gut microbiome through the first 2–3 years of life, after which the gut microbiome reaches a state of high microbial richness equivalent to that of an adult. Thus far, the focus of research has been on the development of gut bacteria. However, the gut ecosystem also comprises viruses,
archaea and eukaryotes such as fungi and protists, whose role in early gut ecosystem development has received less attention.

Both vertical transmission (during pregnancy and birth) and horizontal transmission (from environmental sources like infant diet) have been described as sources of infant gut microbiota. Recent studies provide increasing support for the maternal gut bacterial reservoir as a key source of microbial transmission from mothers to infants. Although the transmission of infectious viruses such as cytomegalovirus and herpes simplex virus in the context of maternal and infant morbidity has been established, little is known about the transmission of bacteriophages (bacterial viruses) from maternal to infant gut. This is partially due to difficulties in isolating and annotating metaviromes. As environmental studies have clearly demonstrated that bacteriophages are key players in the modulation of bacterial communities, it is crucial to study them in the context of the developing human gut ecosystem as the bacterial community is established in the months following birth.

A recent study in 20 healthy infants using viral-like particle (VLP) data provided substantial functional evidence that the bacteriophages colonising the infant gut arise from excisions from pioneering infant gut bacteria. However, the origin of these pioneering bacteriophages, as well as their hosts and their possible roots in the maternal gut ecosystem, remain underexplored. We therefore sequenced 183 total metagenomes (TMs) obtained by isolating all microbial DNA from stool and 66 viral metagenomes (VMs) using a VLP enrichment isolation protocol from 30 mother–infant pairs for whom we have longitudinal samples collected during pregnancy and the first 3 months of life. We first described the bacterial and viral composition in both mothers and infants and elucidated the degree of inter-individual and intra-individual variability. We then compared bacterial strains and viruses across and within mother–infant pairs and found that infant bacterial strains resemble
their mother’s strains more than those of unrelated mothers. We found that the proportion of shared active viruses within mother–infant pairs was lower than the proportion of shared bacteria. However, we also found significantly more viruses that were shared between mother–infant pairs when taking into account bacteriophages incorporated in the bacterial genomes of maternal TMs, which indicates that maternal prophages are a source of pioneering infant gut bacteriophages.

Results

Study sample and gut bacterial composition in mothers and infants

We first characterised the gut bacteriome of mothers and infants and described the degree of sharedness of bacterial species in 183 TMs from 30 mothers and 32 infants, including 2 twin pairs (Figure 1a). Mothers collected their faeces at months 3 and 7 during pregnancy, at birth and at 3 months after birth (Figure 1a). Faecal samples from infants were collected at months 1, 2 and 3 after birth. Although meconium samples collected at birth were available, only 3 of the 32 meconium samples had a sufficient number of microbial reads, reflecting a low bacterial abundance in these samples, and these 3 meconium samples were therefore excluded from the analysis. In this study, mothers had a median age of 32.0±4.4 years, a median pre-pregnancy of BMI of 23.0±2.6 kg/m² (approximately 5 years prior to the pregnancy) and a median gestational age of 39.6±1 week. The infants were 56% female and had a median birth weight of 3,705±380.4 grams. 87.5% of the infants were vaginally delivered, with one twin pair having different delivery modes, and 28% were born at home. In terms of feeding, 53.6% were exclusively breastfed at month 1, 56.3% were exclusively breastfed at month 2 and 53.3% were exclusively breastfed at month 3 (Supplementary Table S1).
The infant gut bacteriome showed a remarkably lower species-level alpha-diversity compared to the maternal gut bacteriome (mean mother=2.62, mean infant=1.26, p-value=1.12e-34, (Supplementary Fig. 1a)), consistent with previous findings\textsuperscript{5}. Over the entire study period, maternal alpha-diversity did not show any significant association patterns to maternal phenotypes or timepoint (Supplementary Fig. 1c), although this analysis may have been hampered by low sample size. The alpha-diversity of bacterial species in infants also did not change significantly during the first 3 months (Supplementary Fig. 1b). At the phylum level, mothers and infants showed remarkably different compositions (Figure 1b, c), with infants having more Actinobacteria and Proteobacteria and mothers having more Bacteroidetes and Firmicutes (FDR<0.05) (Supplementary Table S2).

The species-level bacteriome composition was significantly different in mothers and infants (permutational multivariate analysis of variance (PERMANOVA) p-value=0.001, 999 permutations, Figure 1d). The maternal bacteriome showed greater inter-individual variation than intra-individual variation across timepoints (Supplementary Fig. 2b). Infants also demonstrated more inter-individual variation than intra-individual variation (Supplementary Fig. 2a). We next explored the relation between host phenotypes and the abundance of microbial species and for example found that infants who were exclusively breastfed had a significantly higher abundance of \textit{Cutibacterium acnes}, a bacterium typically known to be a member of skin flora (FDR=0.001) (Supplementary Table S3).
Figure 1: Study description and differences and sharedness in the bacterial composition between mothers and infants.
Maternal to infant vertical transmission of bacteria

Although microbiome composition differed between mothers and infants, 43.6±10.0% of bacterial species were shared across related mother–infant pairs (considering all timepoints in mothers and infants, Figure 1e, Supplementary Table S4). We next aimed to explore whether bacterial species shared between related mother–infant pairs are represented by the same strains. For 12 of the 17 species for which strain-level data could be constructed for both mothers and infants, we found at least one incidence of a mother–infant pair sharing an identical dominant strain. These species were \textit{Akkermansia muciniphila, Alistipes finegoldii, Bacteroides dorei, Bacteroides vulgatus, Bifidobacterium adolescentis, Bifidobacterium bifidum, Bifidobacterium longum, Bifidobacterium breve, Bilophila wadsworthia, Collinsella aerofaciens, Parabacteroides distasonis} and \textit{Bifidobacterium uniformis}. In 10 of these species (B. dorei, B. longum, A. finegoldii, B. uniformis, B. vulgatus, B. adolescentis, B. bifidum, B. wadsworthia, C. aerofaciens and P. distasonis), the phylogenetic distances between related mothers and infants were significantly lower than the distances between unrelated individuals (999 permutations, FDR< 0.05, Figure 2c), suggesting either horizontal or vertical strain transmission of these species from mothers to infants.

We hypothesised that birth mode and delivery location might affect dominant strain-sharing between mothers and infants. The phylogenetic distances of strains of \textit{B. bifidum}
showed a different pattern of transmission from mothers to infants depending on the place of delivery (PERMANOVA, 999 permutations, p=0.001). For home deliveries (28% of deliveries), the B. bifidum phylogenetic distances between infants and their mothers were small, whereas the B. bifidum phylogenetic distances between infants and mothers tended to be larger for hospital deliveries (Figure 2a). Interestingly, in family 20, twin pairs shared identical dominant strains of B. uniformis and B. breve despite having opposing modes of delivery. Furthermore, we observed an interesting pattern in one family where the strain of A. muciniphila showed a strain replacement during pregnancy, with infant strains being similar to those of mother after birth in contrast to maternal strains just before birth (Figure 2b).

Figure 2: Patterns of within- species phylogenetic distances between mothers and infants.
a, pairwise distances between *B. bifidum* strains between mother–infant pairs delivered at the hospital (HS) and mother–infant pairs delivered at home (HM). b, PCoA based on phylogenetic distances in *A. muciniphila* strains in mother-infant pairs. Labels represent timepoints. Only strains present in at least one mother-infant pair are depicted c, Phylogenetic distances within all combinations of samples for related mothers and infants and distances between unrelated individuals.

**Gut viral composition in mothers and infants**

We performed VLP DNA extraction for 93 faecal samples from mothers (pregnancy and birth timepoints) and infants (birth, month 1, month 2 and month 3) that were selected for total metagenome analysis based on sufficiency of faecal material (Figure 1a). In all, 66 VMs were successfully sequenced (Figure 1a) and 51,449 viral-representative contigs were reconstructed from all successfully sequenced VMs, constituting a curated viral database (Figure 3a). Of these, 616 contigs, ranging in size from 3 to 253kbp, had identical ends, which suggests that they represent complete genomes of viruses with circular or terminally redundant linear genomes.

The total number of contigs with an assigned taxonomy at family level was 14,405 (28% of all detected viral contigs). Viral-representative contigs were assigned to 11 families of eukaryotic and prokaryotic viruses. Consistent with previous observations 99.7% of the taxonomically assigned contigs represented bacteriophages. Among the seven bacteriophage families, *Siphoviridae* represented the largest number of contigs. Based on similarity between viral contigs and CRISPR spacer sequences of microbial genomes, we predicted potential hosts for 630 viral contigs (1.2%; Supplementary Table S5; see Methods). Most infant bacteriophages were linked to bacteria of the phyla Actinobacteria, Firmicutes and Proteobacteria whereas maternal bacteriophages were linked to Firmicutes,
Actinobacteria and Bacteroidetes and a very low fraction were linked to Proteobacteria (Figure 3c).

**Figure 3: Characteristics of the viral representative contigs.**

**a,** Distribution of contigs from the curated database by length, taxonomic family and bacterial host. Blue and pink lines represent the spread of contig length in mothers and infants, respectively. The colour of the dots at the median length values indicates if a family includes eukaryotic (yellow), prokaryotic (chartreuse) or unassigned (black) viruses. **b,** Classification of viral contigs by taxonomy, host and viral family. **c,** Number of representative viral contigs linked to their potential bacterial phyla by CRISPR.

To obtain information about the prevalence and presence of viral-representative contigs in VMs, we aligned their reads to the curated viral database (see Methods). Here, we observed that virome composition at viral family level was mostly consistent across maternal samples, with high prevalence of viral contigs assigned to the bacteriophage families *Siphoviridae,*
Podoviridae, Myoviridae and crAss-like phages that were present in all maternal samples (Figure 4a). In maternal samples, the abundances of these most prevalent viral families did not change significantly over time (FDR>0.05, Figure 4a).

Similar to the maternal samples, the viral families Siphoviridae, Podoviridae and Myoviridae were present in all infant samples, and the crAss-like phages were present in 31 out of 32 infant samples. In infants, the abundances of these viral families did not change significantly with time. While abundances of the Podoviridae and Myoviridae viral families were similar between maternal and infant samples, the abundance of crAss-like phages was significantly higher in maternal samples compared to infant samples (FDR<0.05) and the abundance of the Siphoviridae family was higher in infants compared to mothers (FDR<0.05, Figure 4a). Siphoviridae dominated both maternal and infant viromes, showing a greater abundance variation and a nominally significant decrease with time in infants (p-value=0.04, FDR=0.17).

At the level of viral-representative contigs, the composition and diversity of maternal and infant samples showed remarkable differences. The alpha-diversity of the maternal virome was significantly higher than that of the infant virome (p-value=2.2e-05, Figure 4b) and did not change over time in mothers or infants. Similar to the bacterial community, we observed divergent faecal viral communities between mothers and infants, with maternal and infant samples forming two distinct clusters (PERMANOVA, p-value=0.001, 999 permutations, Figure 4c), although this clustering was not as distinct as for the bacteriome composition (beta-diversity R2 0.62 for bacteriome versus 0.37 for virome). Even though infant samples tended to cluster together in the first principal component (Figure 4c), infant viromes were more individual-specific than maternal viromes (PERMANOVA, p-value<0.001, 999 permutations, Figure 4d). Infant viromes also showed greater changes with time than
maternal viromes (PERMANOVA, p-value<0.001, 999 permutations, BC distances: 0.79±0.14 for infants and 0.56±0.14 for mothers) (Figure 4d), indicating more dynamic viromes in the developing gut ecosystems.

Figure 4: Differences in viral composition between mothers and infants.

a, Dynamics of viral families in mothers and infants. Reads were associated with viral lineages based on the annotation of viral contigs. Only viral families present in at least five samples per timepoint
are shown. b, Difference in alpha-diversity (calculated using the Shannon diversity index) of viral-representative contigs in mothers and infants. c, PCoA based on the beta-diversity of the viruses in viral metagenomes (VMs) calculated using Bray-Curtis dissimilarity matrix. Ellipses show 95% confidence intervals. d, Between-individual and within-individual Bray-Curtis distances for infants and mothers.

In mothers, 105 viral contigs were shared across at least 50% of the samples, which is in contrast to only two viral contigs shared across at least 50% of the infant samples, demonstrating, yet again, the high individual-specificity of the infant gut virome.

In mothers, the number of shared viruses across two timepoints was, on average, 1,060, which comprised 26.0±18.6% of all viruses ever detected in a mother (Supplementary Fig.3). This percentage is comparable with another longitudinal study in adults, where 19.4±13.5% of the total viral pool was shared between two timepoints \(^\text{13}\). In infants, on average, 41 viruses, or 20.9±14.9% of all viruses ever detected in an infant, were shared across two timepoints. These results indicate that the percentage of shared viruses across timepoints in an individual is the same for infants and mothers, even though the viral richness in infants was substantially lower.

**Mother to infant transmission of viruses**

A moderate proportion of all the viruses found in infants was shared with their mothers (11.3±20.0%), which is in contrast to the proportion of bacterial species shared between mothers and infants (43.6±10.0%). The proportion of shared viruses did not significantly correlate to any available infant phenotypes.

Previous studies have shown that the adult gut ecosystem observes “Piggyback-the-Winner” dynamics, a state where bacteriophage lysogeny is increasingly favoured in a state with a higher microbial density \(^\text{11,14,15}\). In contrast, the developing infant gut ecosystem is thought to be characterised by a high frequency of bacterial lysis events induced by
bacteriophages\textsuperscript{16}. We therefore sought to compare the cumulative relative abundance of temperate phages between mothers and infants using the presence of integrase and recombinase genes within each viral contig as a measure of temperate phages. The relative abundance of active temperate phages in infants was significantly higher than in mothers (p-value=7.2e-06, Figure 5a) and decreased with timepoint ($R_{\text{Pearson}}$=-0.38, p-value=0.03, Figure 5b). The difference in the abundances of temperate phages in infant and maternal samples may explain the low proportion of shared viruses in mother–infant pairs, as our VLP extraction method targets actively reproducing viruses and not prophages. We therefore investigated whether the proportion of shared viruses increased when accounting for prophages integrated in bacterial genomes in maternal TMs. To do so, we mapped reads from the TMs to the curated viral database and obtained the relative abundances of bacteriophages found in TMs. The viral richness of both mothers and infants was higher in TMs than in VMs, with 462±2767.4 and 1508.5±2968.3 viruses per sample for VMs and TMs, respectively (p-value=3.0e-4). This was presumably due to the presence of prophages in TMs that are not captured as part of the active fraction in the VMs. As in VMs, in TMs, the viral alpha-diversity of maternal samples was higher than those of infants (p-value<2e-16), and the samples from mothers and infants formed two distinct clusters based on Bray-Curtis dissimilarity (PERMANOVA, p-value<0.001, 999 permutations, Supplementary figure 4a). The composition of the virome was highly consistent between concurrent VMs and TMs (Supplementary figure 4b). Interestingly, the proportion of viruses shared between mothers and their infants in TMs was 37.6±22.5% (Figure 5c), which is much higher than in VMs (on average 27.8% higher, p-value=7.6e-06). The proportion of shared viruses between mothers and infants is thus similar to the proportion of the shared bacterial species when the prophages are taken into account. There was no significant correlation between the proportion of shared bacterial species in
mother–infant pairs and the proportion of shared viruses in VMs (Figure 5d). The proportion of shared bacterial species correlates moderately with the proportion of shared viruses in TMs ($R_{\text{Pearson}}=0.47$, $p=0.006$, Figure 5e), further suggesting that vertically transmitted bacteria could be a source for vertically transmitted bacteriophages. These findings together support the argument that the phages colonising the infant gut are vertically transmitted from the mother by means of vertically transmitted bacteria.
Figure 5: Maternal to infant transmission of viruses

a, Relative abundance of temperate phages in mothers (P7, B) and infants (M1, M2, M3). b, The relative abundance of temperate phages in infants decreasing with time. c, The number of the shared viral-representative contigs between related and unrelated infants and mothers in total metagenome (TM) samples. Number of dots represents the number of timepoints available for each infant and the red dots represent the availability of VM’s for that sample d, Correlation of the proportion of shared bacterial and the proportion of shared viruses in VMs of mother–infant pairs. e,
Correlation of the proportion of shared bacterial species and the proportion of shared viruses in TMs of mother–infant pairs.

**Discussion**

This study describes the extent and nature of microbial transmission from mothers to infants, with a focus on viral transmission. A limitation of our study design is that we did not analyse RNA viruses, given their perceived low abundance in the healthy human gut, although future studies should also investigate this dark aspect of the virome. However, while we acknowledge that our sample size only allows us to describe a limited number of associations with phenotypes and only indicates trends in the co-development of bacteriome and virome, our study also has notable strengths. Our novel approach of complementary investigation of both TMs and VMs in both mothers and infants led to novel findings. In the VLP-specific analysis that selectively explores the active virome fraction in stool, we observe high individual-specificity and lower diversity in infants compared to mothers. The average shared fraction of active virome between infants and their mothers was 11.3%, which agrees with the results of one of the few studies to investigate the sharedness of viruses amongst mothers and their infants, where a low sharing of viruses in comparison with bacterial genera was reported in 28 mother–infant twin pairs. However, in our study, when we took prophages into account, the average proportion of vertically transmitted viruses increased to 37.6% and became comparable to the sharedness of gut bacterial species. The majority of the temperate bacteriophages dominating the infant gut were not detected in maternal metaviromes, but were prevalent in maternal TMs, suggesting that infants partly obtain their virome from their mothers via vertically transmitted bacteria.

Our study also addressed previously unstudied factors for maternal to infant microbial transmission such as the place of deliveries. As home deliveries constitute 12.7% of the total
deliveries (2018) in the Netherlands, and 28% in our samples, we possess the unique opportunity to also reveal trends concerning possible novel associations regarding the effect of place of birth on infant bacterial strains like *Bifidobacterium bifidum* which need to be further investigated in a larger cohort\(^7\). An additional strength of our study was that we did not use amplification techniques during the generation of VMs, which allowed accurate quantification of viruses and led to minimal bias in our estimation and characterisation of double stranded DNA viral families.

In conclusion, we provide evidence for mother-to-child viral and bacterial transmission events at high-resolution and novel insights into the early colonisation of the infant gut ecosystem.
Methods

Study design
The samples for this (pilot) study were obtained from the Lifelines NEXT cohort, a birth cohort designed to study the effects of intrinsic and extrinsic determinants on health and disease in a four-generation design. Lifelines NEXT is embedded within the Lifelines cohort study, a prospective three-generation population-based cohort study recording the health and health-related aspects of 167,729 individuals living in Northern Netherlands. In Lifelines NEXT, we aim to include 1,500 pregnant Lifelines participants and intensively follow them, their partners and their children up to at least 1 year after birth. During the Lifelines NEXT study, biomaterials including maternal and neonatal (cord) blood, placental tissue, faeces, breast milk, nasal swabs and urine are collected from the mother and child at 10 timepoints. The long-term health outcomes of these infants and their parents will be investigated. Furthermore, data on medical, social, lifestyle and environmental factors are collected via questionnaires at 14 different timepoints and via connected devices.

Informed consent
The Lifelines NEXT study was approved by the Ethics Committee of the University Medical Center Groningen, document number METC UMCG METc2015/600. Written informed consent forms were signed by the participants or their parents/legal guardians.

Sample collection
Mothers collected their faeces during pregnancy at months 3 and 7, at birth and 3 months after birth (Figure 1a). Faeces from infants were collected from diapers by their parents at months 1, 2 and 3 of infant age. Parents were asked to freeze the stool samples at home at -20°C within 10 min of stool production. Frozen samples were then collected and transported to the UMCG and stored in a -80°C freezer until extraction of microbial and viral DNA. We collected a total of 217 samples for total microbiome analysis.

Total microbial DNA extraction
Microbial DNA was isolated using the QIAamp Fast DNA Stool Mini Kit (Qiagen, Germany) using the QIAcube (Qiagen) from 0.2–0.5 g faecal material, with a final elution volume of 100 μl. DNA eluates were stored at -20°C.
VLP extraction and nucleic acid extraction

Initially, a subset of 93 of 217 faecal samples was selected for viral-like particle extraction based on the availability of faecal material. This subset of 93 samples included maternal samples from pregnancy and birth and infant samples at birth, month 1, month 2 and month 3. The gut virome was studied using extraction of DNA from VLPs as described in Shkoporov et al. 2018. Briefly, 0.5 g of faecal material was resuspended in 10 ml of SM buffer. Samples were centrifuged at 4,800 rcf for 10 min at 4°C. Supernatant was collected and centrifuged again using the same settings. The supernatant was filtered twice through a 0.45-μm pore polyethersulfone membrane filter. VLPs were concentrated from the filtrate with Polyethylene glycol 8000 (Sigma-Aldrich, Cat#P2139) precipitation overnight and purified with chloroform treatment. The resulting fraction was treated with 8 U of TURBO DNase (Ambion/ThermoFisher Scientific) and 20 U of RNase I (ThermoFisher Scientific) at 37°C for 1 h before inactivating enzymes at 70°C for 10 min. Subsequently, proteinase K (40 μg, Sigma-Aldrich, Cat#2308) and 20 μl of 10% SDS were added to the samples and incubated for 20 min at 56°C. Finally, VLPs were lysed by addition of 100 μl of Phage Lysis Buffer (4.5 M guanidinium isothiocyanate, 44 mM sodium citrate pH 7.0, 0.88% sarkosyl, 0.72% 2-mercaptoethanol) and incubated at 65°C for 10 min. Nucleic acids were extracted twice from lysates using Phenol/Chloroform/Isoamyl Alcohol 25:24:1 (ThermoFisher Scientific) treatment followed by centrifugation at 8,000 g for 5 min at room temperature. The resulting aqueous phase was subjected to the final round of purification using the DNeasy Blood & Tissue Kit (Qiagen) with a final elution volume of 50 μl. Viral DNA was stored at -20°C.

Genomic library preparation and sequencing

Faecal microbial DNA and viral DNA samples were sent to Novogene, China for genomic library preparation and shotgun metagenomics sequencing. Sequencing libraries were prepared using the NEBNext® Ultra™ DNA Library Prep Kit or the NEBNext® Ultra™ II DNA Library Prep Kit, depending on the sample DNA concentration, and sequenced using HiSeq 2000 sequencing with 2 × 150 bp paired-end chemistry (Illumina). On average, 30.76±4.2 million paired-end total metagenome reads and 29.8±4.8 million paired-end viral metagenome reads were generated per sample. Of 217 collected samples for total microbiome analysis, 187 were successfully sequenced. Following this, three meconium samples and one mislabelled sample were excluded. For virome analysis, 66 of the 93 samples
were successfully sequenced. As with bacteriome, meconium samples failed sequencing due to their perceived low viral abundance.

Profiling of total gut microbiome composition

Total metagenome sequencing reads were trimmed and Illumina sequence adaptor sequences removed using KneadData tools (v0.7.4) and an average PHRED quality score of 33. Following trimming, the KneadData-integrated Bowtie2 tool (v2.4.2) was used to remove reads that aligned to the human genome (GRCh37/hg19), and the quality of the processed data was examined using the FastQC toolkit (v0.11.9). Samples with a clean read-depth <5 million were not considered in further analysis. Taxonomic composition of total metagenomes was profiled using the MetaPhlAn3 tool with the MetaPhlAn database of marker genes mpa_v30 and the ChocoPhlAn pan-genome database (201901).

Profiling of gut virome composition

VM sequencing reads were subject to quality trimming. Read mapping to the human (GRCh38.p12) reference genome was performed with KneadData (v0.5.1). On average, 22.1±4.5 million paired-end VLP reads passed quality control. The quality of the raw and clean reads was visualised with FastQC (v0.11.7). Bacterial contamination of VMs was assessed by aligning reads to the single copy chaperonin gene cpn60 database. On average, VLP samples contained 3.7±7.9% of bacterial genomic DNA per sample.

We annotated the gut virome composition using the de novo assembly–based method. Whole metagenome de novo assembly was performed per VMs using SPAdes (v3.11.1) in metagenomic mode (-meta) with default settings. On average, 280,363 contigs were assembled for maternal samples and 46,830 contigs were assembled for infant samples. When pooled, 724,612 contigs from the whole dataset were subjected to a redundancy-removal procedure in which contigs with 90% nucleotide identity over 90% of the length of a shorter contig were considered redundant, and the shorter contig was removed. Overall, 274,870 representative pooled contigs larger than 1 kbp were subject to validation as viral.

For the 274,870-representative pooled contigs, we predicted the Open Reading Frames (ORFs) using Prodigal v2.6.3 in metagenomic mode. A Hidden Markov Model (HMM) algorithm (hmmsearch from HMMER v3.2.1 package) was used to compare amino acid sequences of predicted protein products against the HMM database Prokaryotic Virus
Orthologous Groups (pVOGs). Hits were considered significant at an e-value threshold of $10^{-5}$. Ribosomal proteins were identified using a BLASTp search (e-value threshold of $10^{-10}$) against a subset of ribosomal protein sequences from the COG database (release 2014). We used VirSorter v1.0.3 with its expanded built-in database of viral sequences (‘--db 2’ parameter) in the decontamination mode as one of the steps for prediction of viral sequences. Representative contigs larger than 1 kbp were considered viral if they fulfilled at least one of six criteria (similar to those described in [12,13,30]): (1) they produced BLASTn alignments to a viral section of NCBI RefSeq (release 98) with e-values $\leq 10^{-10}$, covering $>90\%$ of contig length at $>50\%$ identity, (2) they had at least three ORFs, producing HMM-hits to the pVOG database with an e-values $\leq 10^{-5}$, with at least two ORFs per 10 kb of contig length, (3) they were VirSorter-positive (all six categories, including suggestive), (4) they were circular, (5) they produced BLASTn alignments to 427 crAss-like reference genomes & unpublished data) with an e-values $\leq 10^{-10}$ covering $>90\%$ of contig length at $>50\%$ identity, or (6) they were longer than 3 kbp with no hits to the nt database (release 235) (alignments $>100$ nucleotides with $90\%$ identity and an e-value of $10^{-10}$). 54,267 contigs fulfilled at least one of these six criteria, and these contigs were subjected to clustering with 427 crAss-like phage contigs from and an unpublished custom database and genomes of the reference database “ProkaryoticViralRefSeq97-Merged”, using vConTACT2 0.9.15 with default parameters. Contigs assigned the status ‘Overlap’, ‘Singleton’ and ‘Outlier’ by vConTACT2 were treated as viral clusters consisting of a single contig in all subsequent analyses. Contigs were further subject to a decontamination procedure based on the following criteria. Contigs were kept if they 1) were circular and contained at least 1 pVOGs hit, 2) were circular and VirSorter-positive, or 3) were VirSorter-positive and did not have ribosomal protein genes. Contigs were excluded if they 1) had more than three ribosomal protein genes (similar to the decontamination process described in [12,13]), or 2) were longer than 200 kbp and co-clustered with contigs possessing more than three ribosomal protein genes.

The final curated database of reconstructed viral sequences generated based on our dataset included 51,455 representative contigs ranging in size from 1 kbp to $>626$ kbp, with low-to-high k-mer coverage (1.1–52,101.3X), which recruited 89% reads per sample on average.
Quality-filtered VM reads were aligned to the custom viral database consisting of 51,455 viral-representative contigs on a per-sample basis using Bowtie2 v2.3.4.1 in ‘end-to-end’ mode. A count table was subsequently generated using SAMTools v1.9. Sequence coverage was calculated per contig per sample using the BEDtools v2.25.0 ‘coverage’ command. To remove spurious Bowtie2 alignments, readcounts that featured a breadth of contig coverage less than 1 × 75% of a contig length were set to zero resulting in 51,449 viral sequences being used for the construction of the final count table. RPKM value transformation was applied to the final count table, and the resulting RPKM count table was used for downstream analysis.

Family-level taxonomic annotations were assigned to viral contigs using the Demovir script (https://github.com/feargalr/Demovir) with default parameters and database. Manual curation of Demovir taxonomy assignment and expansion on unassigned contigs using vConTACT was performed as described previously. Despite using sequencing adaptors binding only dsDNA, a few genomes of ssDNA viruses assigned to the families Circoviridae, Inoviridae, Microviridae and Parvoviridae were reconstructed. These could be either a result of taxonomy misassignment or of catching these viruses during their replication in the duplex form. Given the low bacterial contamination level, we did not exclude the putatively ssDNA virus contigs from the further analysis, assuming that even in case of taxonomy misassignment, these contigs are still of viral origin.

Virome annotation in TMs

Quality-filtered reads from 183 TMs were aligned to the custom viral database consisting of 51,455 viral-representative contigs on a per-sample basis using Bowtie2 v2.3.4.1 in ‘end-to-end’ mode. A count table was generated and transformed as with VMs. The resulting RPKM count table was used to assess the number of shared viruses in related mother-infant pairs. The increase in the number of shared viruses in mother-infant pairs in TMs compared to VMs did not depend on the initial size of the virus pool or any of the infant or maternal phenotypes studied (Spearman correlation, p-value>0.05). All comparisons regarding the VMs and TMs viromes were performed on the 66 samples that had both VM and TM samples.

Prediction of virus hosts based on CRISPR spacer sequences

CRISPR spacer sequences from the CRISPRCas++ database (18.06.2019) were compared to the 51,455 viral contigs using BLASTN 2.7.1+ with the “-task 'blastn-short' -evalue 1e-5”
parameters. Hosts were predicted based on hits characterised by bitscore ≥45 and <2 mismatches, gaps and unaligned spacer nucleotides. Host taxonomy was retrieved from GenBank using BioPerl 1.6.924.

**Ecological measurements and taxonomic comparisons**

To assess bacterial alpha-diversity for each sample, we used species reported in the MetaPhlAn3 output at >0.1% mean relative abundance and present in at least two samples. For the virome alpha-diversity measurements, no filters were applied to the table of RPKM (reads per kilobase per million) counts. The alpha diversity for both bacteriome and virome was calculated using the Shannon diversity index using the `diversity()` function in R package ‘vegan’ v.2.5-7. To test the difference in Shannon diversity index between mother and infant and the effect of timepoint on infant Shannon diversity index, we used a linear model or a linear mixed model, by comparing the suitability of the random effects using the `exactLRT()` function from the ‘RLRsim’ package.

To compare the bacterial phylum abundances of the five most abundant phyla of mothers and infants, we used a linear model or a linear mixed model, by comparing the suitability of the random effects using the `exactLRT()` function from the ‘RLRsim’ package. Prior to analysis, a centred log-ratio (CLR) transformation was performed. FDR correction was applied to correct for multiple testing, with changes considered statistically significant at FDR<0.05 using the Benjamini-Hochberg method.

The Bray-Curtis distances between samples were calculated based on the level of bacterial species and on viral representative contigs. For the Bray-Curtis distances calculation, we used the function `vegdist()` from the R package ‘vegan’ v.2.5-7. To compare the bacterial and viral composition between mothers and infants, PERMANOVA was used with 999 permutations. For these comparisons the aforementioned filters for abundance and presence of bacterial species were used. For the viral-representative contigs-level no filters were applied.

The dynamics of viral family abundances was only assessed families present in at least five individuals per every timepoint (n=4; crAss-like phages, *Myoviridae*, *Podoviridae* and *Siphoviridae*). To normalise the RPKM counts per sample for these viral families, we used a clr transformation. We further excluded outliers if they were not in the range [Q1 – 1.5*IQR; Q3 + 1.15*IQR], where Q1 and Q3 are the 25th and 75th percentiles and IQR is the interquartile
range. Next, to determine if the abundance of the viral families changed over time, we applied a linear (mixed) model that took into account a timepoint (and the repeated measurements). Differentially abundant viral taxa between adults and infants were defined using a linear (mixed) model that took into account the origin of the sample (and the repeated measurements). The choice of model for every viral family was performed by comparing the suitability of the random effects, as described earlier. FDR correction was applied to correct for multiple testing as described previously.

**Bacterial species–specific strain analysis**

Strain SNP haplotypes were generated using StrainPhlAn3. This method is based on reconstructing consensus sequence variants within species-specific marker genes and using them to estimate strain-level phylogenies. This method only takes into account the dominant strain of species and hence our method misses overlaps in secondary strains. We then performed multiple sequence alignment and used the Kimura 2-parameter method from the ‘EMBOSS’ package to calculate phylogenetic distance matrices that contain the pairwise nucleotide substitution rate between strains. Using presence in more than 50 samples as a cut-off, we were able to construct SNP haplotype differences in 17 bacterial species. To identify distinct strain clusters within species, we performed hierarchical clustering using the R function hcluster(). To define identical strains between mother and infant, a definition of 0.0 of Kimura strain distance value was used. The significance of the difference between phylogenetic distances between related mother-infant pairs and unrelated mother-infant pairs was tested using PERMANOVA (999 permutations) and a FDR<0.05 was considered significant. Similarly, the effect of place of birth was tested using PERMANOVA (999 permutations) on the phylogenetic distances of *Bifidobacterium bifidum* correcting for samples origin (mother or infant) and repeated measurements.

**Association with phenotypes**

We used R package ‘Multivariate Association with Linear Models (MaAsLin)’ to associate phenotype data with bacterial species. MaAsLin performs boosted, additive general linear models of associations between phenotypes and the relative abundance of species. We performed this analysis only in infant samples where we used a cut-off of 0.01% species abundance and selected only species occurring in at least 5% of the samples. Prior to input into MaAsLin, we performed a clr transformation. In MaAsLin, we set individual IDs as random
effects and added read-depth, sex, delivery mode, infant feeding type, timepoint, birth weight, place of birth, maternal age, maternal pre-pregnancy BMI and gestational age as fixed effects. We defined statistical significance at FDR<0.05. Given the low sample size to study association of viral contig abundances with phenotypes we did not perform this analysis.

**Code availability**

All codes used in this study can be found at: https://github.com/GRONINGEN-MICROBIOME-CENTRE/Lifelines_NEXT
**List of citations**

1. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).

2. Robertson, R. C., Manges, A. R., Finlay, B. B. & Prendergast, A. J. The Human Microbiome and Child Growth – First 1000 Days and Beyond. *Trends Microbiol.* **27**, 131–147 (2019).

3. Tamburini, S., Shen, N., Wu, H. C. & Clemente, J. C. The microbiome in early life: implications for health outcomes. *Nat. Med.* **22**, 713–722 (2016).

4. Van Daele, E., Knol, J. & Belzer, C. Microbial transmission from mother to child: improving infant intestinal microbiota development by identifying the obstacles. *Crit. Rev. Microbiol.* **45**, 613–648 (2019).

5. Yassour, M. *et al.* Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host Microbe* **24**, 146-154.e4 (2018).

6. Ferretti, P. *et al.* Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe* **24**, 133-145.e5 (2018).

7. Megli, C. J. & Coyne, C. B. Infections at the maternal-fetal interface: an overview of pathogenesis and defence. *Nat. Rev. Microbiol.* 1–16 (2021) doi:10.1038/s41579-021-00610-y.

8. Garmaeva, S. *et al.* Studying the gut virome in the metagenomic era: challenges and perspectives. *BMC Biol.* **17**, 84 (2019).

9. Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W. & Sullivan, M. B. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J.* **10**, 437–449 (2016).

10. Emerson, J. B. *et al.* Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* **3**, 870–880 (2018).

11. Liang, G. *et al.* The stepwise assembly of the neonatal virome is modulated by breastfeeding. *Nature* **581**, 470–474 (2020).
12. Shkoporov, A. N. et al. The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host Microbe* **26**, 527-541.e5 (2019).

13. Garmaeva, S. et al. Stability of the human gut virome and effect of gluten-free diet. *Cell Rep.* **35**, 109132 (2021).

14. Knowles, B. et al. Lytic to temperate switching of viral communities. *Nature* **531**, 466–470 (2016).

15. Redgwell, T. A. et al. Prophages in the infant gut are largely induced, and may be functionally relevant to their hosts. http://biorxiv.org/lookup/doi/10.1101/2021.06.25.449885 (2021) doi:10.1101/2021.06.25.449885.

16. Maqsood, R. et al. Discordant transmission of bacteria and viruses from mothers to babies at birth. *Microbiome* **7**, 156 (2019).

17. De Staat van Volksgezondheid en Zorg. *Bevallingen*. https://www.staatvenz.nl/kerncijfers/bevallingen.

18. Scholtens, S. et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int. J. Epidemiol.* **44**, 1172–1180 (2015).

19. Warmink-Perdijk, W. D. B. et al. Lifelines NEXT: a prospective birth cohort adding the next generation to the three-generation Lifelines cohort study. *Eur. J. Epidemiol.* **35**, 157–168 (2020).

20. Shkoporov, A. N. et al. Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* **6**, 68 (2018).

21. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *elife* **10**, e65088 (2021).

22. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

23. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **7**, 1338 (2018).
24. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

25. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

26. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

27. Grazziotin, A. L., Koonin, E. V. & Kristensen, D. M. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res.* **45**, D491–D498 (2017).

28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

29. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).

30. Clooney, A. G. *et al.* Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. *Cell Host Microbe* **26**, 764-778.e5 (2019).

31. Crits-Christoph, A. *et al.* Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community: Halite metagenome. *Environ. Microbiol.* **18**, 2064–2077 (2016).

32. Guerin, E. *et al.* Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. *Cell Host Microbe* **24**, 653-664.e6 (2018).

33. Bin Jang, H. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).

34. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

35. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
36. Roux, S., Emerson, J. B., Elof-Fadrosh, E. A. & Sullivan, M. B. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **5**, e3817 (2017).

37. Pourcel, C. *et al.* CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. *Nucleic Acids Res.* gkz915 (2019) doi:10.1093/nar/gkz915.

38. Sayers, E. W. *et al.* GenBank. *Nucleic Acids Res.* **49**, D92–D96 (2021).

39. Stajich, J. E. The Bioperl Toolkit: Perl Modules for the Life Sciences. *Genome Res.* **12**, 1611–1618 (2002).

40. Jari Oksanen, M. F., Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O’Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoeecs, Helene Wagner. *vegan: Community Ecology Package*. (2020).

41. Scheipl, F., Greven, S. & Küchenhoff, H. Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comput. Stat. Data Anal.* **52**, 3283–3299 (2008).

42. Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C. & Segata, N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.* **27**, 626–638 (2017).

43. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).

44. Mallick, H. *et al.* Multivariable Association Discovery in Population-scale Meta-omics Studies. *bioRxiv* 2021.01.20.427420 (2021) doi:10.1101/2021.01.20.427420.
Supplementary figures:

**Supplementary Figure 1: Violin plots representing the alpha diversity in mothers and infants**

(a) Difference in alpha-diversity of bacterial species in mothers and infants (*** depicts a p value of <2.2x10e-16). Alpha-diversity of the bacteria at species level was calculated using Shannon diversity index. (b) Alpha-diversity of bacterial species in infants across timepoints. (c) Alpha-diversity of bacterial species in mothers across timepoints.

**Supplementary figure 2: Barplots representing beta-diversity in mothers and infants**

(a) Bray-Curtis within-individual distances between timepoints in infants: Month 1 (M1) to Month 2 (M2), M1 to Month 3 (M3) and M2 to M3. The Bray-Curtis dissimilarity matrix was calculated based on the species level of bacteria. (b) Bray-Curtis within-individual distances in mothers between
timepoints: P3-P7, P3-B, P3-M3, P7-B, P7-M3, B-M3. Bray-Curtis dissimilarity matrix was calculated based on bacterial species.

**Supplementary figure 3: Number of viral-representative contigs per subject present in a given number of timepoints.** Where available (2 infants), three timepoints were subsampled to two timepoints, and the mean was taken into consideration. Y-axis is log-scaled. “1 timepoint” is viral richness for all the infants or mothers at every timepoint.
Supplementary figure 4: Overview of viral contigs in total metagenomes (TMs) and viral metagenomes (VMs).

a, PCoA based on the beta-diversity of the viruses in total metagenomes (TMs) calculated using Bray-Curtis dissimilarity matrix. Ellipses show 95% confidence intervals. b, heatmap showing the similarity between infant individual viromes in VMs and TMs.
Acknowledgements

We thank the participants of the Lifelines NEXT cohort for their collaboration. We thank doctors’ assistants and Service bureau team of Lifelines NEXT, and significant contributions of the persons involved in the collection, processing and storage of samples: Brenda Hijnmans, Annet Jansen, Gea Lamberts, Rianne de Roos, Daphne Teuben and Ettje Tigchelaar. We thank Kate McIntyre for editing the manuscript. We also thank the Genomics Coordination Center for providing data infrastructure and access to high performance computing clusters. We thank Daoming Wang for his assistance in the figure designs.

SG and TS hold scholarships from the Graduate School of Medical Sciences, University of Groningen and the Junior Scientific Masterclass, University of Groningen, respectively. SG was awarded a de Cocks-Hadders Stitching grant (grant number: 2021-08), and TS was awarded de Cocks Hadders Stichting grant (Winston Bakker Fonds WB-08). SB was supported by EUCAN-connect, a federated FAIR platform enabling large-scale analysis of high-value cohort data connecting Europe and Canada in personalized health. Furthermore, this project was funded by the Netherlands Heart Foundation (IN-CONTROL CVON grant 2018-27 to AZ and JF), the Netherlands Organization for Scientific Research (NWO) (NWO Gravitation Exposome-NL (024.004.017) to JF, AK and AZ, NWO-VIDI 864.13.013 and NWO-VICI VI.C.202.022 to JF, NWO-VIDI 016.178.056 to AZ and NWO Spinoza Prize SPI 92-266 to CW), the European Research Council (ERC) (ERC Advanced Grant 2012-322698 to CW, ERC Consolidator Grant 101001678 to JF and ERC Starting Grant 715772 to AZ) and the RuG Investment Agenda Grant Personalized Health to CW. JF and CW are also supported by the Netherlands Organ-on-Chip Initiative, an NWO Gravitation project (024.003.001) funded by the Ministry of Education, Culture and Science of the government of the Netherlands. The study was also financially supported with a public-private partnership allowance of Health Holland Topsector Life Sciences & Health to stimulate public-private partnerships. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Lifelines NEXT cohort study

Jackie Dekens\textsuperscript{1,2}, Aafje Dotinga\textsuperscript{3}, Sanne Gordijn\textsuperscript{4}, Soesma Jankipersadsing\textsuperscript{1}, Ank de Jonge\textsuperscript{5,9}, Marlou L.A. de Kroon\textsuperscript{6}, Gerard H. Koppelman\textsuperscript{7}, Folkert Kuipers\textsuperscript{8}, Lilian L. Peters\textsuperscript{5,9}, Jelmer R. Prins\textsuperscript{4}, Sijmen A. Reijneveld\textsuperscript{6}, Sicco Scherjon\textsuperscript{4}, Jan Sikkema\textsuperscript{2}, Morris A. Swertz\textsuperscript{1}, Henkjan J. Verkade\textsuperscript{8}, Cisca Wijmenga\textsuperscript{1}, Alexandra Zhernakova\textsuperscript{1}

\textsuperscript{1}Department of Genetics, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands
\textsuperscript{2}University Medical Center Groningen, Center for Development and Innovation
\textsuperscript{3}Lifelines cohort study, Groningen, The Netherlands
\textsuperscript{4}Department of Obstetrics and Gynecology, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands
\textsuperscript{5}Department of Midwifery Science, Amsterdam University Medical Centre, Vrije Universiteit Amsterdam, AVAG/Amsterdam Reproduction and Development, Amsterdam, The Netherlands
\textsuperscript{6}Department of Health Sciences, University of Groningen and University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
\textsuperscript{7}Department of Paediatric Pulmonology and Paediatric Allergology, Beatrix Children’s Hospital, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
\textsuperscript{8}Department of Pediatrics, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands
\textsuperscript{9}Department of Midwifery Science, Amsterdam University Medical Center, Vrije Universiteit Amsterdam AVAG/Amsterdam Public Health and Department of General Practice & Elderly Care Medicine, University Medical Center Groningen, section Midwifery Science AVAG, University of Groningen, 9700 RB, Groningen, the Netherlands.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarytablesmotherinfanttransmissionbacterialvirusesNM.xlsx