The human virome: assembly, composition and host interactions

Guanxiang Liang ✉ *and Frederic D. Bushman* ✉

Abstract | The human body hosts vast microbial communities, termed the microbiome. Less well known is the fact that the human body also hosts vast numbers of different viruses, collectively termed the 'virome'. Viruses are believed to be the most abundant and diverse biological entities on our planet, with an estimated 10^{31} particles on Earth. The human virome is similarly vast and complex, consisting of approximately 10^{13} particles per human individual, with great heterogeneity. In recent years, studies of the human virome using metagenomic sequencing and other methods have clarified aspects of human virome diversity at different body sites, the relationships to disease states and mechanisms of establishment of the human virome during early life. Despite increasing focus, it remains the case that the majority of sequence data in a typical virome study remain unidentified, highlighting the extent of unexplored viral 'dark matter'. Nevertheless, it is now clear that viral community states can be associated with adverse outcomes for the human host, whereas other states are characteristic of health. In this Review, we provide an overview of research on the human virome and highlight outstanding recent studies that explore the assembly, composition and dynamics of the human virome as well as host–virome interactions in health and disease.

We are becoming accustomed to the idea that healthy humans are colonized by a rich diversity of microorganisms — the microbiome. However, less well known is that healthy humans are also colonized by a remarkable diversity of viruses — the virome. The human virome comprises bacteriophages (phages) that infect bacteria, viruses that infect other cellular microorganisms such as archaea, viruses that infect human cells and viruses present as transients in food[1–7].

Centuries of medical research have linked infection by specific viruses with characteristic disease states; however, the nature and importance of whole viral populations were mostly not appreciated until the development of advanced DNA sequencing methods that could report the structures of whole communities. Untargeted sequencing of purified viral samples, termed 'shotgun sequencing', was first applied to environmental viral populations in 2002 by Breitbart et al. Viral particles were prepared from seawater, and then shotgun metagenomic sequencing was employed to characterize the viral communities present[8], revealing highly abundant and diverse phage genomes, as well as a large proportion of viral 'dark matter' (that is, sequences that looked like nothing in available databases). The next year, the same research group carried out the first study of whole viral communities from human faeces[1], again revealing rich and diverse viral populations, and emphasizing how little of this diversity had been studied previously. Since then, similar methods have been applied in many studies of virome populations, providing a rich picture of the associations of the human virome with health and disease, and continuing to emphasize the expanse of viral dark matter (reviewed in REFS[9,10]).

The discovery of so much dark matter is not surprising, given that seawater has ~10^7 virus-like particles (VLPs) per millilitre and faeces ~10^9 VLPs per gram. In these studies, particles that look like viral particles are not commonly verified as replication competent; thus, the term VLP is used to reflect the fact that we are uncertain that these particles are replication-competent viruses, although for many it seems likely. These vast populations, inferred to be mostly unstudied phages, are extremely diverse in the number of types as well as overall numbers of particles. Furthermore, in a few cases, viral lineages have been shown to evolve rapidly, also contributing to the observed rich sequence variation[4]. The National Center for Biotechnology Information (NCBI) Genomes database contains only 10,462 complete viral genome sequences (as of February 2021), a tiny fraction of the global diversity. Thus, studying the virome is exciting for the extent of the novelty in every experiment, but also daunting for the analytical challenges.

Viral populations vary greatly across the human body. The human gut contains the most abundant populations,

Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

✉*e-mail: liangguanxiang@gmail.com; bushman@pennmedicine.upenn.edu*

https://doi.org/10.1038/s41579-021-00536-5

Table 1 | **Examples of viral population alterations in human disorders**

Human disease	Sample	Major virome alteration	Ref.
Severe acute malnutrition	Faeces	Reduced viral diversity	102
Crohn's disease and ulcerative colitis	Faeces	Increased *Caudovirales* richness	14
Crohn's disease	Faeces and biopsies	Moderate alterations	152
AIDS	Faeces	Increased enteric adenoviruses	30
Type 1 diabetes	Faeces	Reduced viral diversity	16
Hypertension	Faeces	*Erwinia* phage ΦEaH2 and *Lactococcus* phage 1706 may be associated with hypertension	18
Type 2 diabetes	Faeces	Increased putative phage scaffolds	153
DOCK8 deficiency	Skin swabs	Increased skin virome, especially human papillomavirus	71
Colorectal cancer	Faeces	Increased viral diversity	19
Crohn's disease and ulcerative colitis	Faeces	Increased *Caudovirales* abundance	11
Type 1 diabetes during pregnancy	Faeces	Increased picobirnaviruses and tobamoviruses	17
Bacterial vaginosis	Vaginal swabs	Viral population structures correlated with bacterial vaginosis	74
Early-diagnosed Crohn's disease and ulcerative colitis	Gut biopsies	Increased *Hepadnaviridae* and *Hepeviridae*; reduced *Polydnaviridae*, *Tymoviridae* and *Virgaviridae*	154
Coeliac disease autoimmunity	Faeces	Increased enteroviruses	145
Crohn's disease	Faeces	The virulent phages are replaced with temperate phages	13
Ulcerative colitis	Gut biopsies	Increased *Caudovirales*, phage and bacteria virulence functions, and loss of viral–bacterial correlations	15
HIV viraemia	Seminal fluid	Increased human cytomegalovirus	75
Very early-onset inflammatory bowel disease	Faeces	Increased ratio of *Caudovirales* to *Microviridae*	12
Haematopoietic stem cell transplantation	Faeces	Increased picobirnaviruses	155

and these have been the most studied. This site is rich in cells of both the human gut and prokaryotic microbiota, providing a rich variety of hosts for viral growth. Most other sites have sparser microbiota, at least in healthy individuals, and so sparser viral populations as well; however, recent work has defined rich viral populations at many locations throughout the human body. A high degree of inter-individual variation is seen in the human virome, paralleling findings from bacterial and fungal communities, raising questions of to what extent inter-individual differences in phenotypes are attributable to differences in the virome (TABLE 1).

Recent studies have uncovered numerous factors that show associations with virome structure in addition to anatomical location, such as diet, age and geographic location of the individual sampled. Disease is another prominent influence, with emerging studies suggesting associations between virome structure and inflammatory bowel disease, diabetes, hypertension and cancer[11–19]. Most studies have only reported associations of the virome and diseases — thus, more work is needed to understand the directions of causal relationships.

In this Review, we provide an overview of research on the human virome, focusing both on secure conclusions and on areas where additional work would be helpful. We first discuss the virome at different body sites in healthy humans. Next, we summarize nascent work on the origin of these populations, that is, the initial assembly of the human virome in neonates and infants after delivery. We then summarize data on interactions between the host and the virome, including links to disease states.

Human virome diversity

The number of bacterial cells associated with the human body is estimated to be roughly the same as the number of human cells, ~10^{13} in total[20]. One estimate suggests that each gram of metazoan tissue is associated, on average, with 3×10^9 bacteria[21], and the complexity of these communities increases with body size[22]. VLP counts in humans reveal that the ratios between viruses and bacteria range from roughly 0.1 to 10, suggesting that the total number of viruses in the human body is of a similar order to the number of bacterial and human cells (reviewed in REF.[10]).

Viruses that are found in humans can be categorized by various features. Viral genomes can be either RNA or DNA, and either double-stranded or single-stranded. Genome sizes can be as low as a few kilobases or as large as hundreds of kilobases. All viruses have a protein shell (or capsid) surrounding the genome; viruses can also be enclosed in one or more lipid membranes. Viral particles can be classified according to their morphology — spherical (usually icosahedral), filamentous, bullet-shaped, pleomorphic or, for phages, tailed.

Most constituents of the human virome are inferred to be phages. This is an inference because, in most cases, the majority of sequences uncovered in a virome metagenomic sequencing experiment do not align with any information present in existing databases (BOX 1), so it is unknown whether they are phages or some other virus types. The major taxa of phages are typically *Caudovirales* (tailed phages) and *Microviridae* (icosahedral non-tailed phages) of the group including phage ΦX174. Phages can have multiple relationships with their bacterial hosts, which constrain ideas about virome dynamics (FIG. 1). Lytic phages inject their nucleic acid into cells, leading to the synthesis of new viral components, the assembly of particles and the lysis of host cells, thus releasing progeny phages. Another mode of replication, carried out by temperate phages, can spare the host initially. Phages inject DNA into cells, after which the DNA can then become integrated into the host cell chromosome, forming a prophage. Prophages are maintained in a quiescent state by repressor proteins[23]. However, should conditions become unfavourable in the host cell — for example, by damage to DNA — phage DNA can become excised from the cellular chromosome leading to lytic growth, resulting in cell lysis and the production of viral progeny. Phages can also have other relationships with their hosts — one is pseudolysogeny, in which phage genomes persist as episomes without

Temperate phages
Phages that are able to grow via both lytic and lysogenic replication pathways.

integration. In another mode, phages can bud out of infected cells, sparing the host cell from lysis and death.

Viruses that infect human cells are also an important part of the human virome. Some may cause acute infections, and others may establish long-term latency. Some viruses engage in benign colonization and cannot be associated with any particular disease, appearing to be long-term 'passengers' or 'commensals'. Virome sequencing studies have unveiled some new lineages of human viruses that appear to be such commensals. For example, the *Anelloviridae* are a family of non-enveloped, single-stranded DNA viruses with quite small circular genomes (2–4 kb), including torque teno virus, torque teno mini virus and torque teno midi virus[24–26]. The genomes of viruses in this family encode what appears to be a single large protein, although gene expression has not been

Box 1 | **Methods for studying the virome and the 'dark matter'**

Isolating virus-like particles (VLPs) from samples of interest is the first step of a virome study. Focusing on VLPs avoids wasting resources on sequencing of non-viral nucleic acids. Filtration is a typical first step to purify VLPs, taking advantage of the fact that viruses are smaller than cells. Commonly used filter pore sizes include 0.2 µm and 0.45 µm. Filtering VLPs using 0.2-µm filters may lose large viruses[158], whereas 0.45-µm filters may fail to retain some bacteria, as rare bacteria are <0.45 µm[159]. However, these bacteria appear to be uncommon in humans[160,161], so use of 0.45-µm filters seems optimal. Unpublished comparisons of the two filter types in our laboratory showed no differences in contamination with bacterial 16S rRNA gene sequences, supporting use of 0.45-µm filters.

Other fractionation steps can provide purer VLPs, but often at the expense of losing viral material. Chloroform extraction can be used to disrupt cell membranes, but results in the loss of membrane-enclosed viruses. Comparison of VLP samples purified from human stool with or without chloroform treatment showed that chloroform modestly reduced the recovery of contaminating bacterial DNA. For stool, membrane-enclosed viruses are rare[51], so chloroform treatment may allow superior VLP purification.

For virome nucleic acid preparations, absolute yields are typically low, so multiple displacement amplification using phage Φ29 DNA polymerase has been used to increase DNA amounts. However, this can result in unequal amplification of different DNA forms, with small circular single-stranded DNA genomes amplifying preferentially[162], distorting the ratios of different viral lineages[163]. Nevertheless, this is often the method of choice.

Numerous algorithms, databases and pipelines have been developed to process virome sequencing data, supporting quality control, read dereplication, contig assembly, viral identification, dark matter exploration and other steps (see the table). Despite steadily improving methods, it remains challenging to identify unstudied viruses in complex mixtures. Researchers can align individual sequence reads to databases, allowing identification of many candidate viruses, but attributions are often tentative and a large proportion of reads are typically not attributed to anything (that is, the 'dark matter'). Assembling reads into contigs can provide more opportunity to explore previously unstudied sequences, but still a substantial fraction of contigs may find no attribution. Combining contig analysis with cataloguing gene types, and quantifying matches to gene families known to be viral, can help with attribution. Another helpful step can be assessing circularity in contigs, which commonly indicates completion of a metagenomic viral sequence (although this too can arise as an artefact of sequence repetition in a contig assembly).

No standards of dark matter annotation have been established, so the proportion of viral dark matter varies between studies. Values range from 13 to 95% (reviewed in REFS[9,35]). In some cases, a community will be dominated by a well-known virus, driving down the unknown proportion, but commonly the unknown sequences represent a majority. A recent study constructed 33,242 viral species from 32 available public human virome studies, and >32,000 species were predicted to be phages, which is 12-fold higher than the phage number in the RefSeq database[133]. Thus, metagenomic sequencing boosts the discovery of novel viral genomes enormously. The vast majority of these probable viruses have yet to be grown in cell culture, classically a key step in establishing viral taxa, but today viral lineages are commonly specified as sequences before successful culture. As a consequence, for many members of the virome, it is not known what organisms serve as hosts for replication. The field is developing stand-alone virome-derived contig databases that can be updated regularly, such as the Gut Virome Database (GVD)[133], which are useful for updating new discoveries and establishing standards for annotating the dark matter.

RNA viruses of the human virome have generally been less studied. RNA appears to be less stable in samples typically used for metagenomics, so the RNA viruses may be undersampled for technical reasons. Many studies focus on phages, and only a few RNA phages have been reported. Recent studies have highlighted a much greater diversity of RNA phages than previously thought in certain ecological niches[164,165], focusing interest in further studies of RNA viruses.

Another challenge is DNA contamination in low-biomass samples. Studies have found that low levels of environmental DNA are usually present in laboratory reagents and that this DNA can predominate in samples that start with low amounts of nucleic acid[88,166,167]. The yields of nucleic acids from virome preparations are typically modest, so it is important to consider contamination when interpreting virome data. It is essential to analyse multiple blank samples in parallel with experimental samples so as to document the contamination background — commonly, at least a little of this background will annotate as viral, providing considerable challenges in data analysis. Sequence data from negative controls should be uploaded to databases along with virome data as part of the publication process. One useful tool for evaluating virome enrichment in virome sequence data is ViromeQC[168]. Overall, it is important to be mindful of the potential contribution of contamination and to carefully consider the potential impact on interpretation of experimental results.

Software for analysis of the human virome

Name	Function	Ref.
ViromeQC	Detects contamination	168
CD-HIT	Reads clustering	169
grabseqs	Public data downloading	170
VirSorter	Viral contig annotation	171
VirFinder	Viral contig annotation	172
FastViromeExplorer	Viral contig annotation	173
Metavir	Viral contig annotation	174
MetaPhinder	Viral contig annotation	175
ViromeScan	Viral contig annotation	176
vConTACT	Taxonomic classification	177
NCBI Virus	Database	178
UniProt	Database	179
pVOG	Database	180
Pfam	Database	181
Vfam	Database	182
Sunbeam	Analysis pipeline	183
Cenote-Taker	Analysis pipeline	184
VirusSeeker	Analysis pipeline	185
PHACT	Phage life cycle classification	186
PHASTER	Prophage prediction	187

Viral contigs
Contiguous sequences assembled from overlapping sequence reads, which are then annotated as whole or partial viral genomes.

well studied because representative *Anelloviridae* have not yet been grown in pure culture. *Anelloviridae* are extremely diverse and can be found in many human body sites in a large fraction of all humans examined. No specific pathogenic effects have been linked to *Anelloviridae* so far (reviewed in REFS[27,28]). Greater abundance of *Anelloviridae* has been found in individuals who are immunocompromised, including the recipients of lung transplants, individuals who are HIV positive and individuals on immunosuppressive medications owing to inflammatory bowel disease, indicating that *Anelloviridae* are normally under host immune control[12,14,29–31]. The newly discovered *Redondoviridae* are another family of small circular DNA viruses that appear to be widespread commensals commonly found in the respiratory tract[32–34].

High inter-individual variation in the human virome has been reported in many studies (reviewed in REFS[9,10,35–38]). However, within a healthy adult, the virome is usually relatively stable over time, paralleling stability in the cellular microbiome. For example, one study found that ~80% of viral contigs present persisted over a span of 2.5 years in the gut of one individual[6]. Another recent study tracked the gut virome of 10 individuals and found that >90% of recognizable viral contigs persisted in each individual over 1 year[39]. Studies on the oral virome revealed similar stability[40,41]. As discussed below, destabilization of the virome is often associated with disease states.

Virome of different body sites

Numerous recent studies have characterized the human virome at different body sites, revealing rich populations at numerous locations (FIG. 2). Phages are distributed widely across the human body, and different anatomical sites may have quite different phage composition due to the presence of different host bacteria. The distribution of eukaryotic viruses also differs at different body sites. Some notable site-specific features are summarized below.

Gastrointestinal tract. The gastrointestinal tract is commonly the most abundant site of viral colonization, reaching ~10^9 VLPs per gram of intestinal contents. Analysis of virome sequence data suggests that phages are the most abundant identifiable members of this population (reviewed in REFS[9,10,35–38]). Visualization of

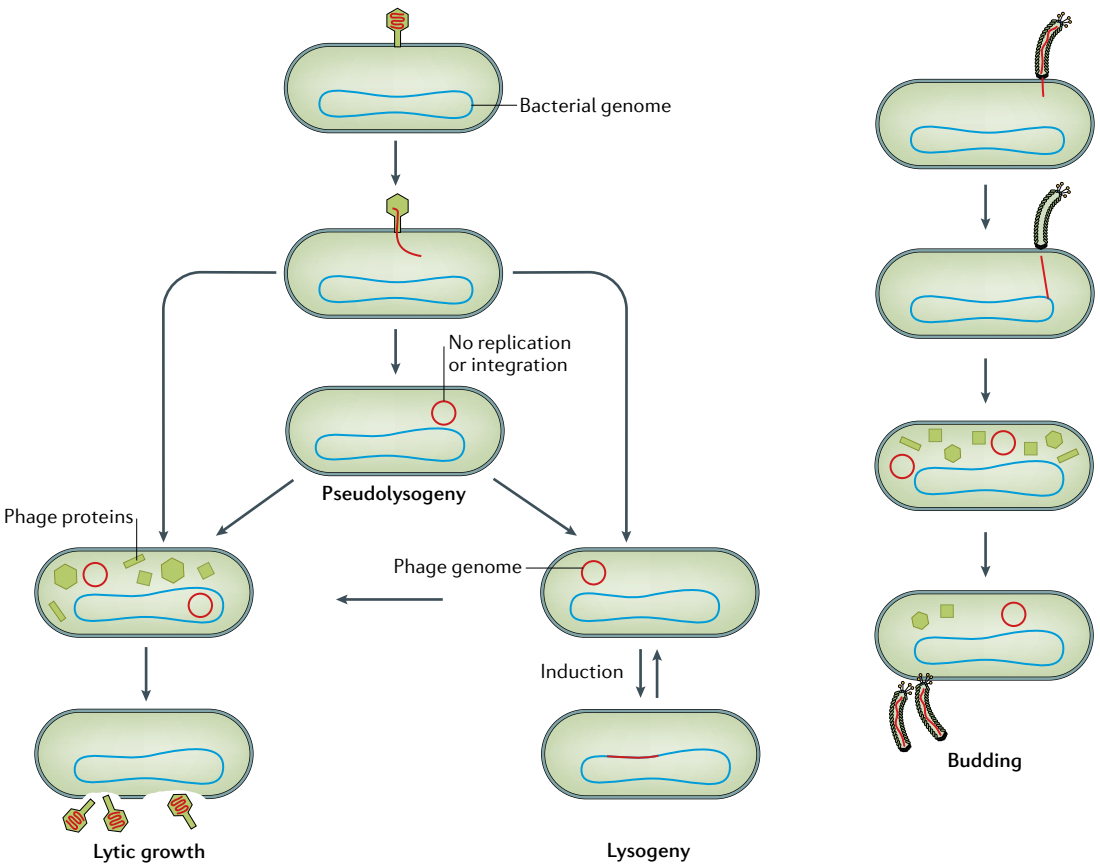

Fig. 1 | **Phage replication cycles.** Phages can engage in four types of interactions with their hosts. In lytic growth, phages infect cells, produce viral macromolecules, assemble new particles and lyse host cells, thereby liberating new viral particles. In lysogenic growth, phages inject their genomes into cells, and the genomes then become integrated into the bacterial cellular chromosome. The prophage is maintained in a quiescent state until detection of a suitable induction signal, after which the phage genome becomes excised and goes on to direct lytic growth. Pseudolysogeny is a loose interaction between the phage and the host in which the phage genome is present in the bacterial cell but not actively directing lytic growth[156]. Lastly, some phages such as the filamentous phages (*Inoviridae*) can infect cells and preserve the infected cell while producing new phage progeny by budding.

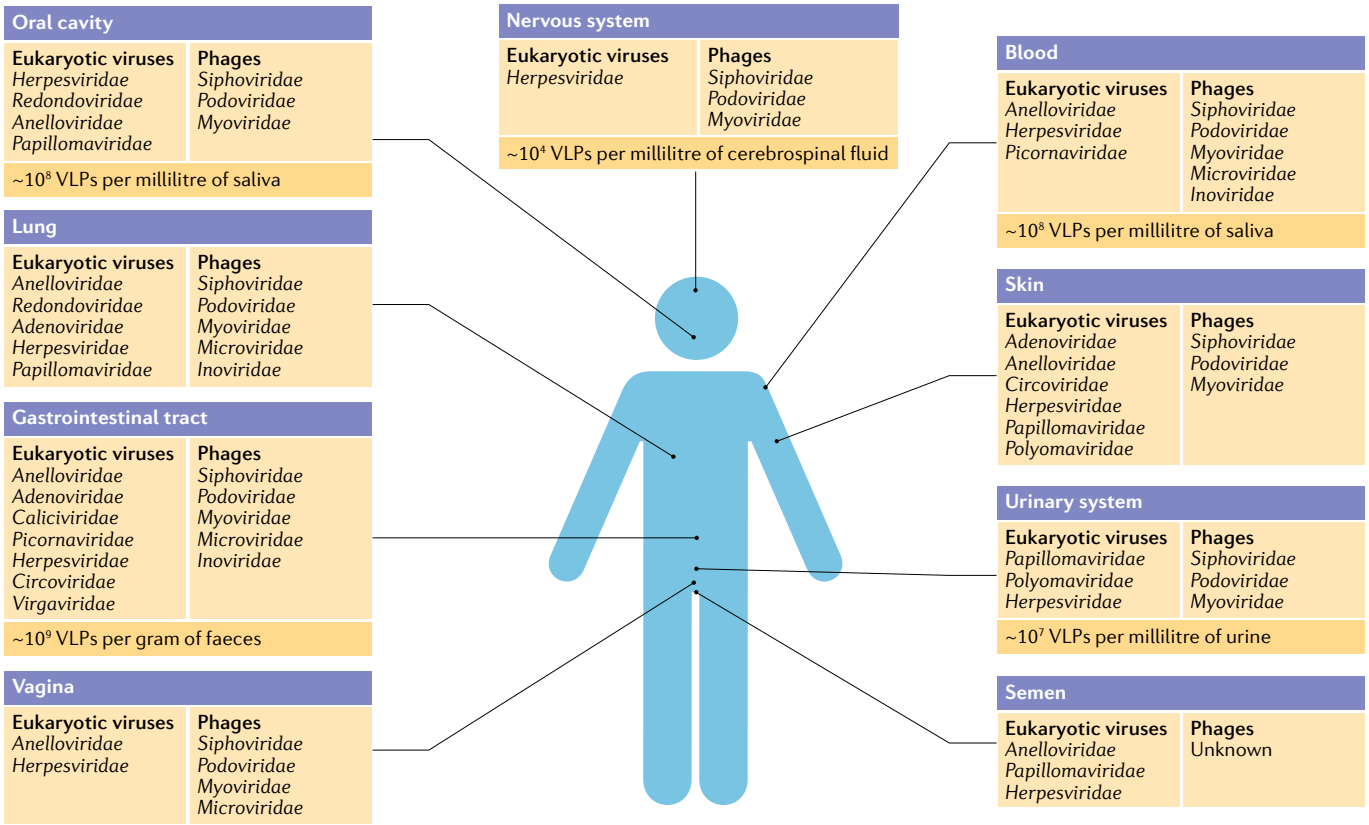

Fig. 2 | **The human virome at different body sites.** Summary of viruses found at each human body site. Viral types are summarized from published virome surveys[2,9,10,29,32–38,40–48,53–74,76–78]; those known at each body site are likely to increase as more human populations are surveyed and more viral types are discovered. In a few cases, viral lineages were excluded from the analysis because they likely represent contaminants or misattributions. These exclusions include mimivirus, phycodnavirus, marseillevirus, flaviviruses and poxviruses in blood, and baculovirus in the vagina. VLP, virus-like particle.

stool VLPs using electron microscopy reveals that the majority of the phages belong to the order *Caudovirales* (tailed phages; reviewed in REF.[10]). Metagenomic sequencing of the human gut virome also indicated that *Caudovirales* is commonly predominant, along with the spherical *Microviridae* (reviewed in REFS[9,10,35–38]).

It has been suggested that the most prevalent phage lineage in the human gut is often crAssphage (cross-assembly phage; members of the *Caudovirales*), specifically the short-tailed podoviruses, which infect bacteria of the phylum Bacteriodetes, common members of the gut microbiota[42–46]. crAssphages are commonly found in greater than 50% of human gut content samples and show a global distribution[42–44]. The abundance of crAssphages can be up to 90% of a human gut viral community[42]. Recently, the crAssphage ΦCrAss001 has been shown to infect *Bacteroides intestinalis*[46], specifying one host species experimentally. Genomic analysis shows that phage genes important for lysogeny are rare or absent in crAssphage genomes[45,46], and a lytic mode of replication has been suggested by infection studies in vitro[46]. However, curiously, proliferation of crAssphages does not appear to reduce the growth rate of their host cells in vitro[46]. One explanation is that crAssphages replicate via pseudolysogeny, in which the phage genome persists in a quiescent state as an episome (reviewed in REF.[47]), only rarely lysing the host cell.

The healthy human gut usually contains relatively low proportions of eukaryotic viruses. DNA viral lineages occasionally detected include *Anelloviridae*, *Geminiviridae*, *Herpesviridae*, *Nanoviridae*, *Papillomaviridae*, *Parvoviridae*, *Polyomaviridae*, *Adenoviridae* and *Circoviridae* (reviewed in REF.[48]). The most commonly detected RNA viruses include *Caliciviridae*, *Picornaviridae*, *Reoviridae* and some plant viruses that appear to originate in food, such as *Virgaviridae*[2,12,49,50].

A notable pattern in viruses of the gut is that most residents, including both phages and human viruses, are not enveloped. This makes sense — lipid envelopes are unlikely to survive the detergent effects of bile salts, dehydration in the large intestine and the conditions of the environment outside the gut required for transmission via the faecal–oral route. A recent statistical analysis showed a significant association between faecal–oral transmission and the absence of a lipid envelope[51]. It will be of interest to investigate virus structure–transmission relationships in more detail and at additional human body sites.

Surprisingly, coronaviruses are an exception; despite possessing a lipid envelope, many coronaviruses are known to be transmitted via the faecal–oral route[51]. For severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), viral RNA has been widely reported in faeces, although the infectious potential of these viruses

Multiple displacement amplification
A whole-genome amplification method, which starts by binding random primers to the template DNA and is then followed by strand displacement DNA synthesis performed by DNA polymerase, usually Φ29 DNA polymerase.

Primary immunodeficiencies
A group of rare immune disorders caused by genetic defects.

is uncertain[52]. One possible explanation is that coronaviruses such as SARS-CoV-2 can replicate in host cells of the lower gastrointestinal tract, so they do not need to traverse the entire gastrointestinal tract in order to appear in faeces; another possibility is that coronavirus particles are relatively stable for enveloped viruses. It will be useful to investigate these questions and clarify the infectivity of coronaviruses in faecal material systematically.

Oral cavity. The human oral cavity contains diverse viral communities as well as complex microbial populations. To date, saliva samples have been the primary source of material to characterize the oral virome[40,53], revealing abundant viral populations. Additional oral microenvironments, such as dental plaque[54], have also been studied, revealing high diversity in these environments as well. Staining of particles with a fluorescent dye that binds DNA, followed by visualization under a fluorescent microscope, shows approximately 10^8 VLPs per millilitre of saliva in healthy humans[55]. The most abundant taxon of phages in the oral virome is the *Caudovirales*[40,41,56].

Common eukaryotic viruses in the oral cavity of healthy adults include *Herpesviridae*, *Papillomaviridae*, *Anelloviridae* and *Redondoviridae*[32,57]. *Anelloviridae* are the most common, but, surprisingly, the newly discovered *Redondoviridae*[32] is the second most common virus family. However, it should be noted that prevalence measures are expected to be dependent on the sampling methods used, and the multiple displacement amplification steps typically used to amplify virome samples likely favour amplification of small DNA circles, potentially boosting detection of *Anelloviridae* and *Redondoviridae*. Reports so far put the prevalence of *Redondoviridae* at 2–15% in different populations[32–34]. As with *Anelloviridae*, whether *Redondoviridae* can cause disease is unknown. Preliminary studies show increased levels of redondoviruses in humans with periodontitis, patients who are critically ill in a medical intensive care unit and individuals with severe respiratory diseases[32,33], although to date there is no evidence that redondoviruses are contributing to the disease states. Thus, both anelloviruses and redondoviruses appear to be common commensal viruses that might be undetected if not for developments in viral metagenomic sequencing methods.

Respiratory tract. Virome analyses have been performed on respiratory tract samples including sputum, nasopharyngeal swabs and bronchoalveolar lavage, showing that the healthy human lung and respiratory tract can be populated by large viral communities[29,58–62]. Among human DNA viruses, *Anelloviridae* has been reported to be the most prevalent family of DNA viruses, followed by *Redondoviridae*[32]. Additional eukaryotic viruses frequently detected include *Adenoviridae*, *Herpesviridae* and *Papillomaviridae*[29,58–62]. Phages are commonly found, including *Caudovirales*, *Microviridae* and *Inoviridae*[29,58–62]. Phages, like most of the cellular microbiota, even when found in the lung appear to be derived mainly from the abundant bacterial populations in the mouth and upper respiratory tract.

Blood. Viruses of blood have been studied closely, to understand human health and also to assess the safety of donor blood supplies. A pioneering study found viral particles in blood using electron microscopy and identified genomic sequences related to several eukaryotic viruses, including *Anelloviridae*[63]. Other recent studies have reported *Herpesviridae*, *Marseilleviridae*, *Mimiviridae*, *Phycodnaviridae* and *Picornaviridae* families, with proportions varying with the geographical site sampled[64,65]. Some of these findings illustrate the challenges of virome studies, where samples with low levels of authentic viruses may be prone to contamination (BOX 1). It is important to note that *Marseilleviridae*, *Mimiviridae* and *Phycodnaviridae* are not known to replicate in human cells, and may represent environmental contamination. Phages belonging to the *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Microviridae* and *Inoviridae* families have also been reported in blood, and, similarly, their origin is unclear (reviewed in REFS[48,66,67]). Some of these studies did not perform viral particle enrichment prior to sequencing, so that prophage sequences integrated in bacterial genomes (rather than viral particles) may have been identified. A recent study reported that phage particles may be transported across gut epithelial cell layers by transcytosis, potentially reaching systemic circulation, with unknown consequences[68]. Thus, much remains to be learnt about the blood virome of healthy individuals; however, it does seem likely that at least *Anelloviridae* are common in blood and so must be expected to be present throughout the blood supply.

Skin. Compared with other body sites, the skin has a relatively low microbial biomass, which can, for some samples, make it difficult to distinguish the resident microbiome and virome from various forms of contamination. Metagenomic analysis of skin swabs revealed the presence of multiple eukaryotic virus families including *Polyomaviridae*, *Papillomaviridae* and *Circoviridae*[69]. In a recent DNA virome study with VLP enrichment[70], ~95% of the viral sequences did not match a known viral genome and of the reads that could be assigned, many were associated with *Caudovirales*. This study also reported that the skin of healthy individuals harboured eukaryotic viruses including *Adenoviridae*, *Anelloviridae*, *Circoviridae*, *Herpesviridae*, *Papillomaviridae* and *Polyomaviridae*[70]. A notable expansion of these eukaryotic viruses was observed on the skin of individuals with primary immunodeficiencies, indicating the importance of immune surveillance in controlling viral colonization of the skin[71].

Urogenital system. Urine samples from healthy humans have been reported to contain viruses in the region of 10^7 VLPs per millilitre[72]. Most of the identifiable viruses were phages; in addition, human papillomaviruses could be identified in >90% of subjects in some cohorts[72,73]. Virome analyses of healthy vaginal samples showed that the majority of identified viral sequences are derived from double-stranded DNA phages, with eukaryotic viruses contributing only 4% of total reads[74]. In seminal fluid, *Anelloviridae*, *Herpesviridae* and multiple genotypes of *Papillomaviridae* have been detected[75]. Thus, for these

Meconium
The first stool of a neonate.

body sites we again see mixtures of viruses that replicate in human cells and viruses from the resident microbiota.

Nervous system. Little information is available on virome populations in the nervous system in healthy humans. A recent study estimated the VLP number at ~10^4 per millilitre of cerebrospinal fluid, with phages predominant, including *Myoviridae*, *Siphoviridae* and *Podoviridae*[76]. *Herpesviridae* were also detectable[76]. The clinical consequences of infection by herpesviruses in the nervous system have been well studied. Herpes simplex viruses, human cytomegalovirus and varicella zoster virus can establish latent infections in the central nervous system without symptoms[77,78]; these viruses can later be reactivated and produce viral particles[78].

To summarize the virome over multiple human body sites, the human gut contains the most abundant viruses and has been the most frequently studied. Lower particle numbers are found at other body sites, but all seem to have detectable viral colonization. For sites with a resident microbiota, the viruses found are typically a mixture of viruses replicating in the local human cells and viruses infecting the local microbiota. The extent of circulation between sites is just starting to be assessed. Upper respiratory microbiota likely contribute much of the microbiome and virome to lower respiratory sites, at least for the phage population. Small circular DNA viruses (*Anelloviridae* and *Redondoviridae*) are common in respiratory samples and can also be found in faecal samples, although it is not known whether they appear in the gut because of replication in the gastrointestinal tract or swallowing of saliva. *Anelloviridae* but not *Redondoviridae* appear in blood, indicating systemic circulation. Even sites that are mostly cut off from normal microbial colonization, such as cerebrospinal fluid, show low levels of viruses including phages. How much of this colonization is due to true local viral replication, how much is due to systemic circulation of viruses and how much is due to technical mishaps associated with reagent contamination remains to be fully clarified. At all of these sites, unannotated 'dark matter' sequences are prominent, emphasizing how much remains to be learnt.

Establishment of the human virome

Timing of the first microbial colonization. The timing of virome establishment is linked to the question of establishment of the microbiome as a whole in human neonates and infants. Historically, the inability to culture microorganisms from samples from healthy deliveries has supported the idea that neonates are usually born sterile. Starting in 2014, several studies using metagenomic sequencing were carried out, leading to the proposal of a microbiome in the placenta, amniotic fluid and even the fetus, implying that microbial colonization may start in utero[79–83]. However, multiple recent studies have indicated that these detections of microorganisms are likely to be false positives due to experimental contamination, and that no placental microbiome is present before rupture of membranes and delivery[84–88]. This raises the question of when viral colonization takes place in human neonates. Vertical transmission of pathogenic viruses during pregnancy has been well documented for rubella virus, human cytomegalovirus, herpes simplex viruses, HIV, Zika virus and human papillomavirus (reviewed in REFS[89–92]). However, these are characteristic of disease states and not health. Virome populations are robust in adults, so the question is when does the virome become established in healthy neonates.

The virome at delivery. An early study of virome colonization sampled meconium shortly after delivery, and failed to find VLPs using epifluorescent microscopy but did report ~10^8 VLPs per gram at 1 week of life, suggesting that the neonate lacked a virome at birth but was quickly colonized[93]. In a study of amniotic fluid, no evidence was found for the existence of a virome in healthy pregnancies before delivery[86]. A later study of neonate stool samples, investigating both RNA and DNA viromes, taken a median of 2.6 days after birth found high diversity in the gut virome, consistent with rapid colonization after delivery[50]. Another metagenomic study using stool samples collected a median of 37 h after birth again showed high viral diversity, supporting rapid virome acquisition[94]. A more recent study of VLPs sampled a median of 17 h after birth reported that only 15% of samples were positive[49]. Thus, the picture that emerges is that neonates usually lack a detectable virome at birth, but are rapidly colonized after rupture of membranes and delivery (FIG. 3).

The first detectable viruses — a predominance of phages. Several studies have investigated the nature of the human virome in samples taken very early in life. Recognizable early colonizers were mainly phages of the *Siphoviridae*, *Podoviridae* and *Myoviridae* families. Another phage family, the *Microviridae*, are less abundant during early life but rise in abundance with age[49,50]. Early bacterial colonizers commonly include *Escherichia*, *Klebsiella*, *Enterococcus*, *Staphylococcus* and *Streptococcus* species[95,96], and the phages of these bacteria are some of the most common early virome members.

A recent study investigated the production of the virome in gut samples of infants at 1 month of life and concluded that lytic phages were relatively rare[49]. Ongoing replication of lytic phages could not be detected in direct infection assays; viral sequences were annotated primarily as lysogenic phages and not lytic phages; and *Microviridae* and crAssphages, which usually do not form lysogens, were rare or absent during the first month in infant guts. Instead, prophage induction was found to contribute most of the gut virome. Bacterial strains were isolated from infants' stools, and many were found to produce viral particles at high levels. The viruses that were produced could be detected as prophage sequences in the bacterial genome sequences. Sequences of the induced phages were also commonly found in the infant stool virome, and the abundance of each type of VLP in stool was positively correlated with the abundance of the host bacteria in the same sample.

These experiments suggested a high rate of phage induction in the infant gut, raising the question of what constitutes the inducing signal. DNA damage is the most studied signal (reviewed in REF.[97]). In vitro studies suggested that spontaneous induction rates may be

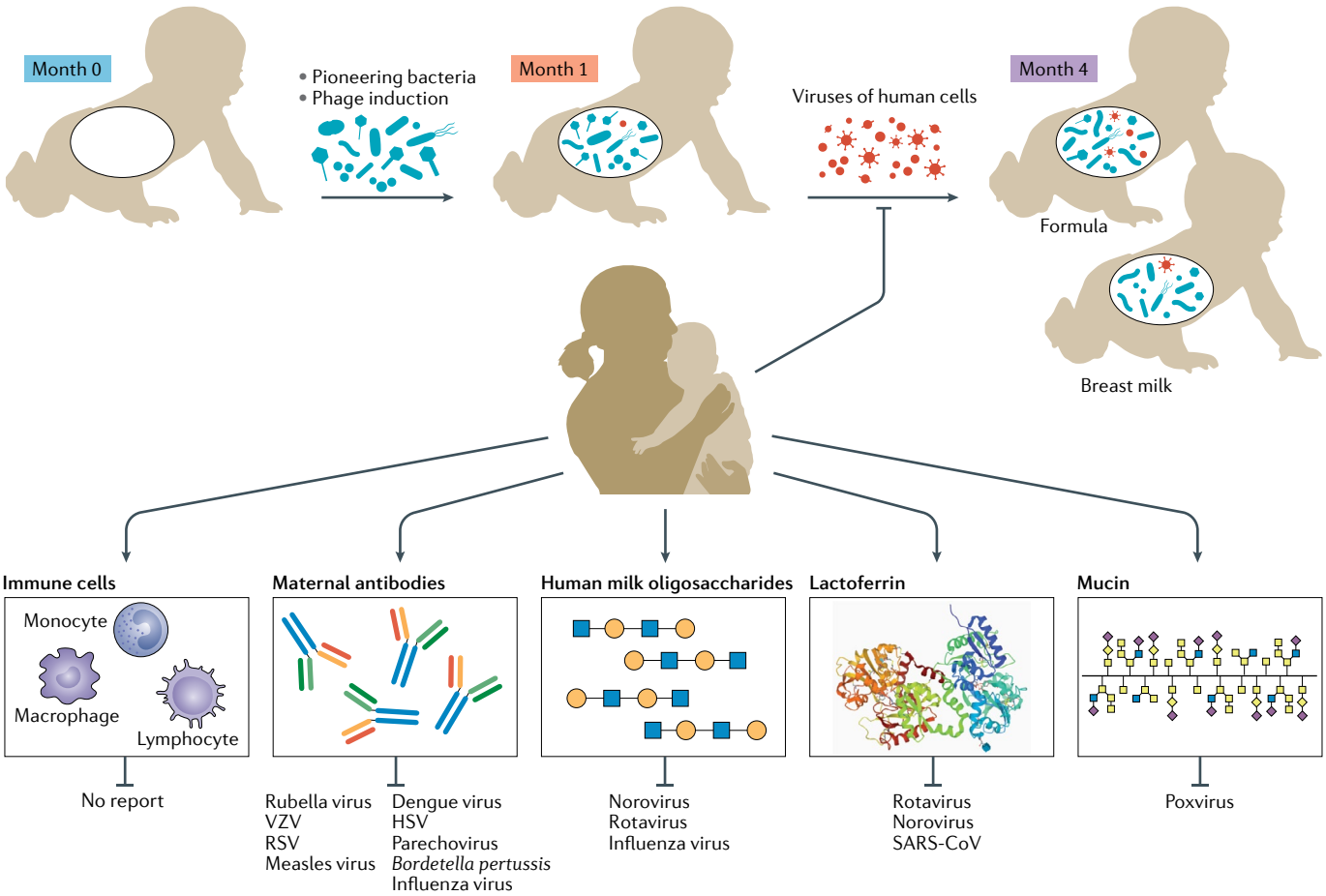

Fig. 3 | **Stepwise assembly of the paediatric virome.** Healthy neonates are typically born lacking a gut virome or microbiome. Pioneering bacteria colonize the gut of the neonate, such that infants have a detectable microbiome by month 1 of life. These bacteria commonly harbour integrated prophages, which occasionally induce prophages, providing a first wave of viral particles in the gut. Later, by month 4, more viruses that infect human cells can be detected. Infection with these viruses, some of which can be pathogenic, is inhibited by breastfeeding. Breastfeeding can also alter the types of phages present by altering the proportions of bacteria in the infant gut, which consequently alters the proportions of their phages. The protective effects of breast milk can be conferred by maternal immune cells or various macromolecules. These include maternal antibodies, human milk oligosaccharides, lactoferrin, mucin and gangliosides. The viral groups targeted by each antiviral factor in human breast milk are shown. Supporting references include REFS[49,111,113–121,157]. HSV, herpes simplex viruses; RSV, respiratory syncytial virus; SARS-CoV; severe acute respiratory syndrome coronavirus; VZV, varicella zoster virus.

relatively low[98,99]. Bacterial metabolites, nutrients and bile salts have all been shown to induce prophages in some models[100,101]. Thus, the signals (if there are any) are unclear and an interesting topic for future studies.

Later evolution of the paediatric virome. The paediatric virome continues to mature with age. Lytic phages appear to become more common later in life. For example, the crAssphages are mostly absent in the first month of life but become more prominent by month 4 (REF.[49]). The gut virome also often undergoes a shift from *Caudovirales*-dominated to *Microviridae*-dominated[12,50,102].

Another question is the possible influence of the mode of delivery. A metagenomic analysis of faecal virome samples from 20 infants at 1 year of age compared spontaneous vaginal delivery with caesarean section and found that the birth mode resulted in distinctly different viral communities, with infants born by spontaneous vaginal delivery having greater viral diversity[103]. The effect of the delivery mode has been reproduced in some, but not all, cohorts in other studies[49]. Studies linking the birth mode and long-term health outcomes reported greater risk for various diseases, including asthma, obesity and diabetes, in children delivered by caesarean section, although the results are controversial and research is ongoing[104–108]. It may be useful to consider possible influences of the virome as well.

Colonization of infants with viruses infecting human cells. The viruses that replicate in human cells are also detected in metagenomic surveys of samples taken in early life. Gastroenteritis is one of the leading causes of childhood mortality, resulting in more than two million deaths every year[109], and viral pathogens are primary causes (reviewed in REF.[110]), highlighting the importance of paediatric virome studies. Viruses that have been reported to be associated with childhood diarrhoea include rotavirus, astrovirus, calicivirus, picornavirus,

polyomavirus and adenovirus (reviewed in REF.[110]). Less well known is the fact that these viruses are commonly found in healthy infant guts in metagenomic studies[49,50]. Other common viruses in healthy infant guts include parvovirus and anellovirus[49,50].

Thus, recent data suggest that healthy infants are colonized in a stepwise fashion. In the first step, prophage induction from pioneering bacteria provides an initial population. Later, lytic phages become more common, and also viruses that replicate in human cells.

Factors that shape the human virome

Numerous factors have been reported to influence the human virome and, ultimately, affect health (FIG. 4), starting in infancy and extending throughout the life of the individual.

Diet. The infant virome, including both phages and eukaryotic viruses, can be affected by diet. Breastfeeding is well established to reduce viral gastroenteritis and infant mortality (reviewed in REFS[111,112]). A recent metagenomic virome study showed lower accumulation of animal cell viruses in guts of infants fed with breast milk[49] (FIG. 3). The protective effect was seen in cohorts from both the United States and Botswana[49]. Viruses affected included *Adenoviridae*, *Picornaviridae* and *Caliciviridae*. Breast milk contains multiple components that protect children from intestinal infections, such as maternal antibodies, oligosaccharides and lactoferrin (reviewed in REF.[111]). These antiviral components have been reported to inhibit viruses, such as rotavirus, norovirus, enterovirus, influenza virus and SARS-CoV (reviewed in REFS[111,113–121]). The phage population structure in infant stool can also be influenced by breastfeeding — specific bacteria are known to increase in abundance with breastfeeding, and in a recent report their phages were increased in abundance as well[49]. Alterations of the viral population in infant stool samples owing to breastfeeding has been proposed in another study, although the small sample size was a limitation[94]. Furthermore, it has been suggested that the infant virome may, in fact, be at least partially transmitted from the mother via breast milk[122,123].

Diet has also been reported to affect the virome in adults. For example, comparison of the virome structure in adult subjects on two different controlled diets showed that individuals on the same diet showed more similar viral compositions than those on different diets[6].

Host genetics and immunity. Intense interest has focused on the potential influence of human genetics on the microbiome, raising the linked question of the role of human genetics in programming the virome. Several studies of twins, comparing monozygotic and dizygotic twin pairs, have suggested an influence of human genetics on microbiome composition because the monozygotic twins showed more similarity[124–126]. By contrast, a recent large-scale study of healthy adults (non-twins) reported little effect of genetics and emphasized environmental factors[127]. In studies of the infant virome, the gut virome compositions of co-twins were more similar than those between unrelated individuals[50,94,102], but this similarity was not strongly affected by zygosity[94,102], emphasizing the importance of the shared environment over human genetic composition. In an early study of adult twins, no greater similarity of virome samples was seen in twin pairs[3]. In a more recent study with a larger cohort of adult monozygotic twins, some were found to have greater similarity in virome composition compared with unrelated individuals, but in other co-twin pairs the microbiota and the virome had diverged to the degree that the twins were no longer notably similar[128]. Collectively, studies of the gut virome in twin pairs so far show similarities early in life, but not stronger similarities in monozygotic twins versus dizygotic twins, thus emphasizing the importance of shared environmental factors in colonization versus contributions of genetic make-up.

However, in inherited diseases such as primary immunodeficiencies, the effects of genetics on the virome are well established. In some cases, phenotypes of mutations in human genes can only be understood by considering the interaction with the virome. One drastic example involves epidermodysplasia verruciformis (also known as treeman syndrome, which is a rare heritable skin disease). In the presence of a mutation in *TMC6* (*EVER1*) or *TMC8* (*EVER2*), skin papillomaviruses replicate aggressively and cause massive inappropriate outgrowth of skin cells, resulting in treeman syndrome (reviewed in REF.[129]). Other primary immunodeficiencies may similarly allow specific viruses or microorganisms to replicate unchecked, resulting in distinctive disorders. The effects of immunodeficiencies on the skin virome were mentioned above; also, virome studies of individuals undergoing treatment for X-linked severe combined

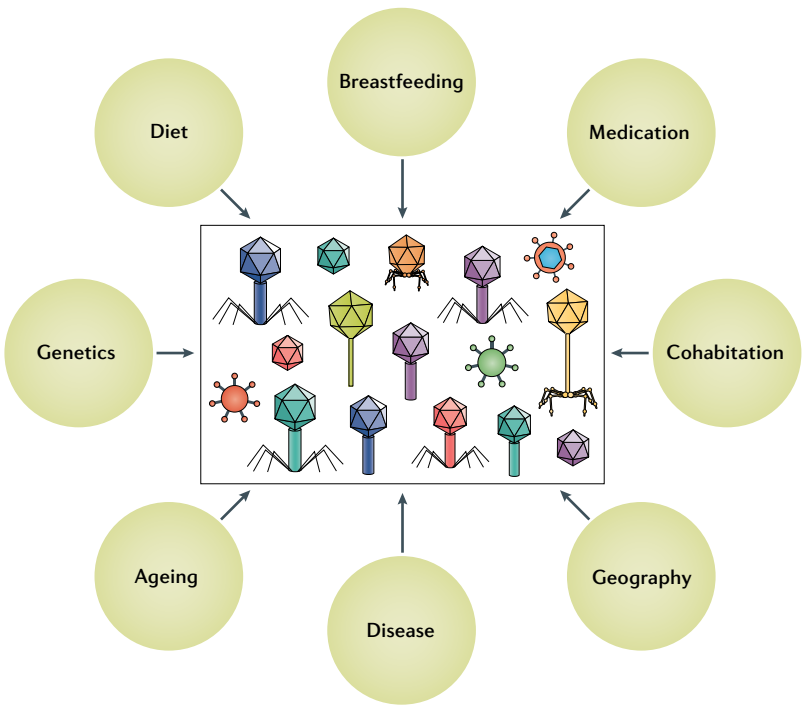

Fig. 4 | **Factors that shape the human virome.** Major factors include the diet[6], breast milk or formula feeding[49,94], medications (including antibiotics and immunosuppressants)[12,131], host genetics[3,50,94,102,128,130], cohabitation[53], geography[49,65,131–134], presence of disease (TABLE 1) and ageing[133].

immunodeficiency (SCID-X1) revealed distinctive outgrowths of several viral lineages in the gut[130]. An interesting question for future investigation is how much of the phenotypes of diverse human genetic diseases are a result of altered interactions with the normal virome.

Geography and stochastics of colonization. Large-scale virome studies have provided evidence that geographic location and stochastics of colonization have strong impacts on human virome variation. A study of human faecal samples from different regions within China reported variation of the phage population structure and found that geography had the strongest impact compared with other variables, including diet, ethnicity and medication[131]. Geography has been associated with the eukaryotic virus populations as well. An early study found geographic variation in eukaryotic viromes in children with diarrhoea from two locations within Australia, with differential prevalence of *Adenoviridae* and *Picornaviridae*[132]. A blood virome analysis of eukaryotic viruses in a Chinese population revealed a distinctive pattern for people living in the southern part of China[65]. Moreover, one study of the infant virome found a higher prevalence of viruses infecting human cells in cohorts of people of African descent than in cohorts from the United States[49]. A recent study collected public metagenomic data sets and built a virome database with >30,000 viral genomes, and found higher viral diversity in populations from non-western countries compared with populations from western countries[133]. Similar results were reported in another study[134]. Thus, effects of geography are quite prominent.

Additional factors. Additional factors have been tested and reported to influence the human virome. One study tested age using public data, and found that viral diversities in early life and in older individuals (>65 years of age) are lower than those in healthy adults (18–65 years of age), indicating another dimension of age-dependent patterns[133]. Effects of ethnicity and medication were also found in a large Chinese cohort[131]. Cohabitation is also a factor — in an early study, members of the same household shared more similarity in the oral virome compared with those from different households[53], suggesting transmission of the virome via close contact.

Thus, nascent studies of factors affecting the virome highlight diet, geography, age and health status as major correlates of virome structure. Genetics is less clear, at least in studies so far of healthy twins. More studies of large cohorts with linked comprehensive metadata will be helpful going forward.

The virome in health and disease

Virome populations can influence their human hosts in numerous ways. Eukaryotic viruses that infect human cells establish infections, trigger immune responses and, sometimes, cause diseases. Phages can affect the host indirectly via modulating of bacterial composition and bacterial fitness. Some phages and human cell viruses can be integrated into cells of their respective hosts, on occasion transferring new functionality to host cells. Some phages may also interact with human cells directly and trigger immune responses (FIG. 5). Recent examples of host–virome interactions in health and disease are discussed below (for reviews, see REFS[135,136]).

Interactions between bacteria, phage and their human hosts. Relatively little is known about the impact of phage predation on human bacterial communities. A window on phage predation and host health is provided by phage therapy, in which phages are deliberately applied to human individuals to treat bacterial infections. As more drug-resistant bacteria are emerging, there is increased interest in this approach (reviewed in REF.[137]). Recently, several studies have used phage cocktails to treat bacterial infection in a few individuals, with apparent success (reviewed in REF.[138]) motivating larger clinical trials. Evidence that the phages are in fact creating evolutionary pressure on pathogenic bacteria in these studies comes from the observation that bacteria often mutated to become resistant to the therapeutic phage. Phages may even be useful to treat additional diseases. For example, a phage cocktail targeting adherent invasive *Escherichia coli* strains has been suggested recently as a treatment for Crohn's disease[139].

Phages can move DNA between cells and, thereby, introduce new functionality to bacterial genomes, which may modify bacterial fitness and virulence (reviewed in REF.[140]). In a recent study using a mouse model, the stool virome was analysed after antibiotic treatment, showing that there was an enrichment of phage-encoded genes for antibiotic resistance[141], which increased resistance in the bacterial community.

Prophage induction can also lead to bacterial cell lysis, regulating bacterial abundance. A recent publication investigated diet-mediated prophage induction of human-associated bacterial strains in vitro, revealing that several common food compounds can inhibit bacterial growth by inducing prophages, including artificial sweeteners[142].

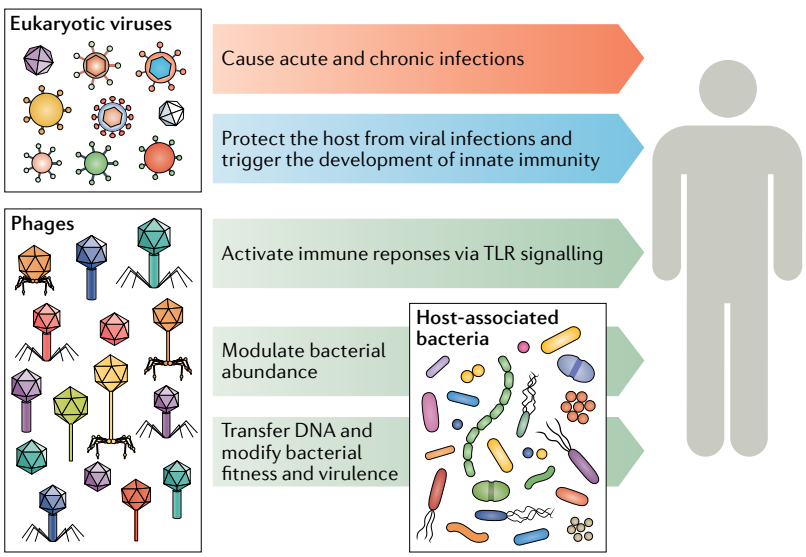

Fig. 5 | **Host–virome interactions.** Eukaryotic viruses have both detrimental (red arrow) and beneficial (blue arrow) effects on host health. Phages interact with the host directly or indirectly via the host-associated bacterial community and pose undetermined effects (green arrow) on host health. Data based on REFS[68,137–144,149–151]. TLR, Toll-like receptor.

Recent in vitro and animal studies indicated that phages may interact with the host immune system directly. Phages may be taken up by immune cells and trigger immune responses, without the mediation of bacteria, via Toll-like receptor (TLR) signalling. A recent publication reported that phages produced by pathogenic bacteria can be taken up by immune cells (dendritic cells, B cells and monocytes) in mice to induce type I interferon responses via TLR3 signalling[143]. Another study showed that interferon-γ (IFNγ)-producing $CD4^+$ T cells and $CD8^+$ T cells were increased in mucosal sites in germ-free mice fed an *E. coli* phage isolated from the human gut[144]. Furthermore, this study also revealed that *Lactobacillus*, *Escherichia* and *Bacteroides* phages can stimulate the production of IL-12, IL-6, IL-10 and IFNγ via the nucleotide-sensing receptor TLR9 (REF.[144]). Collectively, the interactions among phages, bacteria and the host immune system likely have important roles in host immune homeostasis.

Human disease-associated virome signatures. Studies of interactions between the virome with human diseases are just starting. Of course, numerous individual viruses are well known to cause morbidity and mortality. Studies of whole viral populations have now also begun to show patterns associated with disease states. Some recent examples are described below.

The virome has been considered to be a potential trigger of autoimmune diseases. In one study, changes in viral populations were both directly and inversely associated with the development of paediatric type 1 diabetes[16]. Virome signatures have been associated with paediatric and adult inflammatory bowel disease in several studies[11–15], including a reproducible expansion of *Caudovirales* and a reduction of *Microviridae*. Whether this is a consequence of altered dysbiotic bacterial populations or is more deeply involved in the disease process remains to be clarified. A metagenomic study showed that children with frequent exposure to enterovirus between 1 and 2 years of age have a higher risk of coeliac disease[145]. Recent studies revealed that the relationships between phages and bacterial populations may influence growth stunting in children[146,147]. Examples of proposed associations between diseases and viral population structure are listed in TABLE 1.

Studies using animal models have indicated that some eukaryotic viruses may even be beneficial to the host (reviewed in REFS[135,148]). For example, persistent infection by a strain of murine norovirus can compensate for the absence of bacteria in gnotobiotic mice, allowing restoration of intestinal morphology and promoting lymphocyte differentiation[149]. Murine astrovirus protected immunodeficient mice from enteric norovirus infections through the induction of type III interferons in the intestinal epithelial barrier[150]. Depletion of murine gut viruses using an antiviral cocktail inhibited the development of the intestinal intraepithelial lymphocytes, at least in part[151]. So far, all of these studies showing positive effects were performed in model organisms — it will be of interest to determine how much beneficial immune instruction may be directed by the virome in humans.

Thus, both phages and eukaryotic viruses can promote host health through interactions with the host immune system. However, virome studies using human cohorts suggest that dysbiosis of the virome is associated with multiple diseases. These studies emphasize the balance between beneficial and harmful roles of viral populations in humans and other organisms.

Conclusions and perspectives

The virome field has come a long way since the first paper in 2002 that reported metagenomic sequencing of a viral specimen. The methods for study have advanced, although there are still many challenges associated with working with metagenomic dark matter. Human virome population diversity and composition is being documented for many body sites — one consistent conclusion is that each individual harbours diverse and distinctive viral communities. Future studies are needed to clarify the DNA and RNA viromes at different anatomical sites and to link the alterations of viral composition to specific diseases. We now have a detailed picture of the stepwise nature of the assembly of the human virome after birth. The impact of viral colonization during early life on long-term health outcomes is still unknown and warrants careful study. Breastfeeding appears to have an important role during viral colonization; how the antiviral components in breast milk interact with the different components of the virome warrants further investigation. Emerging data indicate that factors that influence the human microbiome also often influence the virome as well, so that sorting out the influence of each will be a challenge going forward. Resident viruses are not only actively interacting with other microorganisms but also with the mammalian immune system. Many intriguing conclusions have so far only been obtained in animal studies or in vitro experiments, focusing attention on translation to studies in humans. Associations between alterations of virome and disease states are being identified more commonly, but in many cases causality and molecular mechanisms remain to be worked out. The vast world of the human virome is beginning to be understood, laying the ground work for numerous future studies of its importance.

Published online: 30 March 2021

1. Breitbart, M. et al. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **185**, 6220–6223 (2003).
2. Zhang, T. et al. RNA viral community in human feces: prevalence of a pathogenic viruses. *PLoS Biol.* **4**, 0108–0118 (2006).
3. Reyes, A. et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
4. Minot, S. et al. Rapid evolution of the human gut virome. *Proc. Natl Acad. Sci. USA* **110**, 12450–12455 (2013).
5. Minot, S., Wu, G. D., Lewis, J. D. & Bushman, F. D. Conservation of gene cassettes among diverse viruses of the human gut. *PLoS ONE* **7**, e42342 (2012).
6. Minot, S. et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).
7. Minot, S., Grunberg, S., Wu, G. D., Lewis, J. D. & Bushman, F. D. Hypervariable loci in the human gut virome. *Proc. Natl Acad. Sci. USA* **109**, 3962–3966 (2012).
8. Breitbart, M. et al. Genomic analysis of uncultured marine viral communities. *Proc. Natl Acad. Sci. USA* **99**, 14250–14255 (2002).
9. Aggarwala, V., Liang, G. & Bushman, F. D. Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob. DNA* **8**, 12 (2017).
10. Shkoporov, A. N. & Hill, C. Bacteriophages of the human gut: the "known knknown" of the microbiome. *Cell Host Microbe* **25**, 195–209 (2019).

11. Fernandes, M. A. et al. Enteric virome and bacterial microbiota in children with ulcerative colitis and Crohn disease. *J. Pediatr. Gastroenterol. Nutr.* **68**, 30–36 (2019).
12. Liang, G. et al. Dynamics of the stool virome in very early-onset inflammatory bowel disease. *J. Crohn's Colitis* **14**, 1600–1610 (2020).
13. 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).
14. Norman, J. M. et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160**, 447–460 (2015).
15. Zuo, T. et al. Gut mucosal virome alterations in ulcerative colitis. *Gut* **68**, 1169–1179 (2019).
16. Zhao, G. et al. Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. *Proc. Natl Acad. Sci. USA* **114**, E6166–E6175 (2017).
17. Kim, K. W. et al. Distinct gut virome profile of pregnant women with type 1 diabetes in the ENDIA study. *Open Forum Infect. Dis.* **6**, ofz025 (2019).
18. Han, M., Yang, P., Zhong, C. & Ning, K. The human gut virome in hypertension. *Front. Microbiol.* **9**, 3150 (2018).
19. Nakatsu, G. et al. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. *Gastroenterology* **155**, 529–541.e5 (2018).
20. Sender, R., Fuchs, S. & Milo, R. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* **14**, 1–14 (2016).
21. Kieft, T. L. & Simmons, K. A. Allometry of animal–microbe interactions and global census of animal-associated microbes. *Proc. Royal. Soc. B.* **282**, 20150702 (2015).
22. Sherrill-Mix, S. et al. Allometry and ecology of the bilaterian gut microbiome. *mBio* **9**, e00319-18 (2018).
23. Jacob, F., Sussman, R. & Monod, J. On the nature of the repressor ensuring the immunity of lysogenic bacteria [French]. *C. R. Acad. Sci.* **254**, 4214–4216 (1962).
24. Nishizawa, T. et al. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochem. Biophys. Res. Commun.* **241**, 92–97 (1997).
25. Takahashi, K., Iwasa, Y., Hijikata, M. & Mishiro, S. Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. *Arch. Virol.* **145**, 979–993 (2000).
26. Ninomiya, M. et al. Identification and genomic characterization of a novel human torque teno virus of 3.2 kb. *J. Gen. Virol.* **88**, 1939–1944 (2007).
27. Freer, G. et al. The virome and its major component, anellovirus, a convoluted system molding human immune defenses and possibly affecting the development of asthma and respiratory diseases in childhood. *Front. Microbiol.* **9**, 686 (2018).
28. Spandole, S., Cimponeriu, D., Berca, L. M. & Mihăescu, G. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch. Virol.* **160**, 893–908 (2015).
29. Young, J. C. et al. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am. J. Transpl.* **15**, 200–209 (2015).
30. Monaco, C. L. et al. Altered virome and bacterial microbiome in human immunodeficiency virus-associated acquired immunodeficiency syndrome. *Cell Host Microbe* **19**, 311–322 (2016).
31. Li, L. et al. AIDS alters the commensal plasma virome. *J. Virol.* **87**, 10912–10915 (2013).
32. Abbas, A. A. et al. Redondoviridae, a family of small, circular DNA viruses of the human oro-respiratory tract associated with periodontitis and critical Illness. *Cell Host Microbe* **25**, 719–729.e4 (2019).
33. Spezia, P. G. et al. Redondovirus DNA in human respiratory samples. *J. Clin. Virol.* **131**, 104586 (2020).
34. Lázaro-Perona, F. et al. Metagenomic detection of two vientoviruses in a human sputum sample. *Viruses* **12**, 327 (2020).
35. Mirzaei, M. K. & Maurice, C. F. Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nat. Rev. Microbiol.* **15**, 397–408 (2017).
36. Lim, E. S., Wang, D. & Holtz, L. R. The bacterial microbiome and virome milestones of infant development. *Trends Microbiol.* **24**, 801–810 (2016).
37. Virgin, H. W. The virome in mammalian physiology and disease. *Cell* **157**, 142–150 (2014).
38. Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F. & Gordon, J. I. Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat. Rev. Microbiol.* **10**, 607–617 (2012).
39. Shkoporov, A. N. et al. The human gut virome is highly diverse, stable, and individual specific. *Cell Host Microbe* **26**, 527–541.e5 (2019).
40. Abeles, S. R. et al. Human oral viruses are personal, persistent and gender-consistent. *ISME J.* **8**, 1753–1767 (2014).
41. Abeles, S. R., Ly, M., Santiago-Rodriguez, T. M. & Pride, D. T. Effects of long term antibiotic therapy on human oral and fecal viromes. *PLoS ONE* **10**, e0134941 (2015).
42. Dutilh, B. E. et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5**, 1–11 (2014).
43. 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).
44. Edwards, R. A. et al. Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nat. Microbiol.* **4**, 1727–1736 (2019).
45. Yutin, N. et al. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat. Microbiol.* **3**, 38–46 (2018).
46. Shkoporov, A. N. et al. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat. Commun.* **9**, 1–8 (2018).
47. Sutton, T. D. S. & Hill, C. Gut bacteriophage: current understanding and challenges. *Front. Endocrinol.* **10**, 784 (2019).
48. Rascovan, N., Duraisamy, R. & Desnues, C. Metagenomics and the human virome in asymptomatic individuals. *Annu. Rev. Microbiol.* **70**, 125–141 (2016).
49. Liang, G. et al. The stepwise assembly of the neonatal virome is modulated by breastfeeding. *Nature* **581**, 470–474 (2020).
50. Lim, E. S. et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **21**, 1228–1234 (2015).
51. Bushman, F. D., McCormick, K. & Sherrill-Mix, S. Virus structures constrain transmission modes. *Nat. Microbiol.* **4**, 1778–1780 (2019).
52. Zapor, M. Persistent detection and infectious potential of SARS-CoV-2 virus in clinical specimens from COVID-19 patients. *Viruses* **12**, 1384 (2020).
53. Robles-Sikisaka, R. et al. Association between living environment and human oral viral ecology. *ISME J.* **7**, 1710–1724 (2013).
54. Naidu, M., Robles-Sikisaka, R., Abeles, S. R., Boehm, T. K. & Pride, D. T. Characterization of bacteriophage communities and CRISPR profiles from dental plaque. *BMC Microbiol.* **14**, 1–13 (2014).
55. Pride, D. T. et al. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *ISME J.* **6**, 915–926 (2012).
56. Pérez-Brocal, V. & Moya, A. The analysis of the oral DNA virome reveals which viruses are widespread and rare among healthy young adults in Valencia (Spain). *PLoS ONE* **13**, e0191867 (2018).
57. Baker, J. L., Bor, B., Agnello, M., Shi, W. & He, X. Ecology of the oral microbiome: beyond bacteria. *Trends Microbiol.* **25**, 362–374 (2017).
58. Willner, D. et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS ONE* **4**, e7370 (2009).
59. Wylie, K. M., Mihindukulasuriya, K. A., Sodergren, E., Weinstock, G. M. & Storch, G. A. Sequence analysis of the human virome in febrile and afebrile children. *PLoS ONE* **7**, e27735 (2012).
60. Clarke, E. L. et al. Microbial lineages in sarcoidosis a metagenomic analysis tailored for low-microbial content samples. *Am. J. Respir. Crit. Care Med.* **197**, 225–234 (2018).
61. Abbas, A. A. et al. The perioperative lung transplant virome: torque teno viruses are elevated in donor lungs and show divergent dynamics in primary graft dysfunction. *Am. J. Transplant.* **17**, 1313–1324 (2017).
62. Abbas, A. A. et al. Bidirectional transfer of *Anelloviridae* lineages between graft and host during lung transplantation. *Am. J. Transplant.* **19**, 1086–1097 (2019).
63. Breitbart, M. & Rohwer, F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques* **39**, 729–736 (2005).
64. Moustafa, A. et al. The blood DNA virome in 8,000 humans. *PLoS Pathog.* **13**, e1006292 (2017).
65. Liu, S. et al. Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese population history. *Cell* **175**, 347–359.e14 (2018).
66. Castillo, D. J., Rifkin, R. F., Cowan, D. A. & Potgieter, M. The healthy human blood microbiome: fact or fiction? *Front. Cell. Infect. Microbiol.* **9**, 148 (2019).
67. Zárate, S., Taboada, B., Yocupicio-Monroy, M. & Arias, C. F. Human virome. *Arch. Med. Res.* **48**, 701–716 (2017).
68. Nguyen, S. et al. Bacteriophage transcytosis provides a mechanism to cross epithelial cell layers. *mBio* **8**, 1874–1891 (2017).
69. Foulongne, V. et al. Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. *PLoS ONE* **7**, e38499 (2012).
70. Hannigan, G. D. et al. The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome. *mBio* **6**, e01578-15 (2015).
71. Tirosh, O. et al. Expanded skin virome in DOCK8-deficient patients. *Nat. Med.* **24**, 1815–1821 (2018).
72. Santiago-Rodriguez, T. M., Ly, M., Bonilla, N. & Pride, D. T. The human urine virome in association with urinary tract infections. *Front. Microbiol.* **6**, 14 (2015).
73. Garretto, A., Miller-Ensminger, T., Wolfe, A. J. & Putonti, C. Bacteriophages of the lower urinary tract. *Nat. Rev. Urol.* **16**, 422–432 (2019).
74. Jakobsen, R. R. et al. Characterization of the vaginal DNA virome in health and dysbiosis. *Viruses* **12**, 1143 (2020).
75. Li, Y. et al. Semen virome of men with HIV on or off antiretroviral treatment. *AIDS* **34**, 827–832 (2020).
76. Ghose, C. et al. The virome of cerebrospinal fluid: viruses where we once thought there were none. *Front. Microbiol.* **10**, 2061 (2019).
77. Meyding Lamade, U. & Strank, C. Herpesvirus infections of the central nervous system in immunocompromised patients. *Ther. Adv. Neurol. Disord.* **5**, 279–296 (2012).
78. McGavern, D. B. & Kang, S. S. Illuminating viral infections in the nervous system. *Nat. Rev. Immunol.* **11**, 318–329 (2011).
79. Aagaard, K. et al. The placenta harbors a unique microbiome. *Sci. Transl Med.* **6**, 237ra65 (2014).
80. Antony, K. M. et al. The preterm placental microbiome varies in association with excess maternal gestational weight gain. *Am. J. Obstet. Gynecol.* **212**, 653.e1–653.e16 (2015).
81. Prince, A. L. et al. The placental membrane microbiome is altered among subjects with spontaneous preterm birth with and without chorioamnionitis. *Am. J. Obstet. Gynecol.* **214**, 627.e1–627.e16 (2016).
82. Collado, M. C., Rautava, S., Aakko, J., Isolauri, E. & Salminen, S. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Sci. Rep.* **6**, 1–13 (2016).
83. Martinez, K. A. et al. Bacterial DNA is present in the fetal intestine and overlaps with that in the placenta in mice. *PLoS ONE* **13**, e0197439 (2018).
84. Theis, K. R. et al. Does the human placenta delivered at term have a microbiota? Results of cultivation, quantitative real-time PCR, 16S rRNA gene sequencing, and metagenomics. *Am. J. Obs. Gynecol.* **220**, 267.e1–267.e39 (2019).
85. Lauder, A. P. et al. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* **4**, 29 (2016).
86. Lim, E. S., Rodriguez, C. & Holtz, L. R. Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. *Microbiome* **6**, 87 (2018).
87. Leiby, J. S. et al. Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome* **6**, 196 (2018).
88. de Goffau, M. C. et al. Human placenta has no microbiome but can contain potential pathogens. *Nature* **572**, 329–334 (2019).
89. Epps, R. E., Pittelkow, M. R. & Daniel Su, W. P. TORCh syndrome. *Semin. Cutan. Med. Surg.* **14**, 179–186 (1995).
90. Leeper, C. & Lutzkanin, A. Infections during pregnancy. *Prim. Care* **45**, 567–586 (2018).
91. Carlson, A., Norwitz, E. R. & Stiller, R. J. Cytomegalovirus infection in pregnancy: should all women be screened? *Rev. Obstet. Gynecol.* **3**, 172–179 (2010).

92. Arora, N., Sadovsky, Y., Dermody, T. S. & Coyne, C. B. Microbial vertical transmission during human pregnancy. *Cell Host Microbe* **21**, 561–567 (2017).
93. Breitbart, M. et al. Viral diversity and dynamics in an infant gut. *Res. Microbiol.* **159**, 367–373 (2008).
94. Maqsood, R. et al. Discordant transmission of bacteria and viruses from mothers to babies at birth. *Microbiome* **7**, 156 (2019).
95. Bäckhed, F. et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).
96. Baumann-Dudenhoeffer, A. M., D'Souza, A. W., Tarr, P. I., Warner, B. B. & Dantas, G. Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes. *Nat. Med.* **24**, 1822–1829 (2018).
97. Sausset, R., Petit, M. A., Gaboriau-Routhiau, V. & De Paepe, M. New insights into intestinal phages. *Mucosal Immunol.* **13**, 205–215 (2020).
98. Nanda, A. M., Thormann, K. & Frunzke, J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host–microbe interactions. *J. Bacteriol.* **197**, 410–419 (2015).
99. Cortes, M. G., Krog, J. & Balázsi, G. Optimality of the spontaneous prophage induction rate. *J. Theor. Biol.* **483**, 110005 (2019).
100. Jubelin, G. et al. Modulation of enterohaemorrhagic *Escherichia coli* survival and virulence in the human gastrointestinal tract. *Microorganisms* **6**, 115 (2018).
101. De Paepe, M. et al. Carriage of λ latent virus is costly for its bacterial host due to frequent reactivation in monoxenic mouse intestine. *PLOS Genet.* **12**, e1005861 (2016).
102. Reyes, A. et al. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl Acad. Sci. USA* **112**, 11941–11946 (2015).
103. McCann, A. et al. Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ* **6**, e4694 (2018).
104. Black, M., Bhattacharya, S., Philip, S., Norman, J. E. & McLernon, D. J. Planned cesarean delivery at term and adverse outcomes in childhood health. *J. Am. Med. Assoc.* **314**, 2271–2279 (2015).
105. Kuhle, S., Tong, O. S. & Woolcott, C. G. Association between caesarean section and childhood obesity: a systematic review and meta-analysis. *Obes. Rev.* **16**, 295–303 (2015).
106. Adlercreutz, E. H., Wingren, C. J., Vincente, R. P., Merlo, J. & Agardh, D. Perinatal risk factors increase the risk of being affected by both type 1 diabetes and coeliac disease. *Acta Paediatr.* **104**, 178–184 (2015).
107. Rutayisire, E., Huang, K., Liu, Y. & Tao, F. The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. *BMC Gastroenterol.* **16**, 86 (2016).
108. Neu, J. & Rushing, J. Cesarean versus vaginal delivery: long-term infant outcomes and the hygiene hypothesis. *Clin. Perinatol.* **38**, 321–331 (2011).
109. Hug, L., Alexander, M., You, D. & Alkema, L. National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: a systematic analysis. *Lancet Glob. Heal.* **7**, e710–e720 (2019).
110. Oude Munnink, B. B., Hoek, L. & van der Hoek, L. Viruses causing gastroenteritis: the known, the new and those beyond. *Viruses* **8**, 42 (2016).
111. Turin, C. G. & Ochoa, T. J. The role of maternal breast milk in preventing infantile diarrhea in the developing world. *Curr. Trop. Med. Rep.* **1**, 97–105 (2014).
112. Lamberti, L. M., Fischer Walker, C. L., Noiman, A., Victora, C. & Black, R. E. Breastfeeding and the risk for diarrhea morbidity and mortality. *BMC Public Health* **11**, S15 (2011).
113. Wakabayashi, H., Oda, H., Yamauchi, K. & Abe, F. Lactoferrin for prevention of common viral infections. *J. Infect. Chemother.* **20**, 666–671 (2014).
114. Lang, J. et al. Inhibition of SARS pseudovirus cell entry by lactoferrin binding to heparan sulfate proteoglycans. *PLoS ONE* **6**, e23710 (2011).
115. Witkowska-Zimny, M. & Kaminska-El-Hassan, E. Cells of human breast milk. *Cell. Mol. Biol. Lett.* **22**, 11 (2017).
116. Simister, N. E. Placental transport of immunoglobulin G. *Vaccine* **21**, 3365–3369 (2003).
117. Pou, C. et al. The repertoire of maternal anti-viral antibodies in human newborns. *Nat. Med.* **25**, 591–596 (2019).
118. Albrecht, M. & Arck, P. C. Vertically transferred immunity in neonates: mothers, mechanisms and mediators. *Front. Immunol.* **11**, 555 (2020).
119. Wiciński, M., Sawicka, E., Gębalski, J., Kubiak, K. & Malinowski, B. Human milk oligosaccharides: health benefits, potential applications in infant formulas, and pharmacology. *Nutrients* **12**, 266 (2020).
120. Berlutti, F. et al. Antiviral properties of lactoferrin — a natural immunity molecule. *Molecules* **16**, 6992–7012 (2011).
121. Conesa, C. et al. Isolation of lactoferrin from milk of different species: calorimetric and antimicrobial studies. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **150**, 131–139 (2008).
122. Pannaraj, P. S. et al. Shared and distinct features of human milk and infant stool viromes. *Front. Microbiol.* **9**, 1162 (2018).
123. Duranti, S. et al. Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome* **5**, 66 (2017).
124. Goodrich, J. K. et al. Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).
125. Goodrich, J. K. et al. Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* **19**, 731–743 (2016).
126. Xie, H. et al. Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst.* **3**, 572–584.e3 (2016).
127. Rothschild, D. et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
128. Moreno-Gallego, J. L. et al. Virome diversity correlates with intestinal microbiome diversity in adult monozygotic twins. *Cell Host Microbe* **25**, 261–272.e5 (2019).
129. Orth, G. Genetics of epidermodysplasia verruciformis: insights into host defense against papillomaviruses. *Semin. Immunol.* **18**, 362–374 (2006).
130. Clarke, E. L. et al. T cell dynamics and response of the microbiota after gene therapy to treat X-linked severe combined immunodeficiency. *Genome Med.* **10**, 70 (2018).
131. Zuo, T. et al. Human–gut–DNA virome variations across geography, ethnicity, and urbanization. *Cell Host Microbe* **28**, 741–751.e4 (2020).
132. Holtz, L. R. et al. Geographic variation in the eukaryotic virome of human diarrhea. *Virology* **468**, 556–564 (2014).
133. Gregory, A. C. et al. The Gut Virome Database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe* **28**, 724–740.e8 (2020).
134. Rampelli, S. et al. Characterization of the human DNA gut virome across populations with different subsistence strategies and geographical origin. *Environ. Microbiol.* **19**, 4728–4735 (2017).
135. Neil, J. A. & Cadwell, K. The intestinal virome and immunity. *J. Immunol.* **201**, 1615–1624 (2018).
136. Seo, S. U. & Kweon, M. N. Virome–host interactions in intestinal health and disease. *Curr. Opin. Virol.* **37**, 63–71 (2019).
137. Moghadam, M. T. et al. How phages overcome the challenges of drug resistant bacteria in clinical infections. *Infect. Drug Resist.* **13**, 45–61 (2020).
138. Kortright, K. E., Chan, B. K., Koff, J. L. & Turner, P. E. Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. *Cell Host Microbe* **25**, 219–232 (2019).
139. Galtier, M. et al. Bacteriophages targeting adherent invasive *Escherichia coli* strains as a promising new treatment for Crohn's disease. *J. Crohns Colitis* **11**, 840–847 (2017).
140. Taylor, V. L., Fitzpatrick, A. D., Islam, Z. & Maxwell, K. L. The diverse impacts of phage morons on bacterial fitness and virulence. *Adv. Virus Res.* **103**, 1–31 (2019).
141. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222 (2013).
142. Boling, L. et al. Dietary prophage inducers and antimicrobials: toward landscaping the human gut microbiome. *Gut Microbes* **11**, 721–734 (2020).
143. Sweere, J. M. et al. Bacteriophage trigger antiviral immunity and prevent clearance of bacterial infection. *Science* **363**, eaat9691 (2019).
144. Gogokhia, L. et al. Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis. *Cell Host Microbe* **25**, 285–299.e8 (2019).
145. Lindfors, K. et al. Metagenomics of the faecal virome indicate a cumulative effect of enterovirus and gluten amount on the risk of coeliac disease autoimmunity in genetically at risk children: the TEDDY study. *Gut* **69**, 1416–1422 (2019).
146. Khan Mirzaei, M. et al. Bacteriophages isolated from stunted children can regulate gut bacterial communities in an age-specific manner. *Cell Host Microbe* **27**, 199–212.e5 (2020).
147. Desai, C. et al. Growth velocity in children with environmental enteric dysfunction is associated with specific bacterial and viral taxa of the gastrointestinal tract in Malawian children. *PLoS Negl. Trop. Dis.* **14**, e0008387 (2020).
148. Lee, S. & Baldridge, M. T. Viruses RIG up intestinal immunity. *Nat. Immunol.* **20**, 1563–1564 (2019).
149. Kernbauer, E., Ding, Y. & Cadwell, K. An enteric virus can replace the beneficial function of commensal bacteria. *Nature* **516**, 94–98 (2014).
150. Ingle, H. et al. Viral complementation of immunodeficiency confers protection against enteric pathogens via interferon-λ. *Nat. Microbiol.* **4**, 1120–1128 (2019).
151. Liu, L. et al. Commensal viruses maintain intestinal intraepithelial lymphocytes via noncanonical RIG-I signaling. *Nat. Immunol.* **20**, 1681–1691 (2019).
152. Pérez-Brocal, V. et al. Metagenomic analysis of Crohn's disease patients identifies changes in the virome and microbiome related to disease status and therapy, and detects potential interactions and biomarkers. *Inflamm. Bowel Dis.* **21**, 2515–2532 (2015).
153. Ma, Y., You, X., Mai, G., Tokuyasu, T. & Liu, C. A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* **6**, 24 (2018).
154. Ungaro, F. et al. Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease. *Gut Microbes* **10**, 149–158 (2019).
155. Legoff, J. et al. The eukaryotic gut virome in hematopoietic stem cell transplantation: new clues in enteric graft-versus-host disease. *Nat. Med.* **23**, 1080–1085 (2017).
156. Łoś, M. & Wegrzyn, G. in *Advances in Virus Research* Vol. 82 339–349 (Academic Press, 2012).
157. Drew, H. R. et al. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl Acad. Sci. USA* **78**, 2179–2183 (1981).
158. Schulz, F. et al. Giant virus diversity and host interactions through global metagenomics. *Nature* **578**, 432–436 (2020).
159. Wang, Y., Hammes, F., Düggelin, M. & Egli, T. Influence of size, shape, and flexibility on bacterial passage through micropore membrane filters. *Environ. Sci. Technol.* **42**, 6749–6754 (2008).
160. Weil, A. A., Becker, R. L. & Harris, J. B. *Vibrio cholerae* at the intersection of immunity and the microbiome. *mSphere* **4**, e00597-19 (2019).
161. Ryan, M. P. & Pembroke, J. T. *Brevundimonas* spp: emerging global opportunistic pathogens. *Virulence* **9**, 480–493 (2018).
162. Roux, S. et al. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* **2016**, e2777 (2016).
163. Kim, K. H. & Bae, J. W. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Environ. Microbiol.* **77**, 7663–7668 (2011).
164. Krishnamurthy, S. R., Janowski, A. B., Zhao, G., Barouch, D. & Wang, D. Hyperexpansion of RNA bacteriophage diversity. *PLOS Biol.* **14**, e1002409 (2016).
165. Callanan, J. et al. Expansion of known ssRNA phage genomes: from tens to over a thousand. *Sci. Adv.* **6**, eaay5981 (2020).
166. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
167. Kim, D. et al. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome* **5**, 1–14 (2017).
168. Zolfo, M. et al. Detecting contamination in viromes using ViromeQC. *Nat. Biotechnol.* **37**, 1408–1412 (2019).
169. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
170. Taylor, L. J., Abbas, A. & Bushman, F. D. grabseqs: simple downloading of reads and metadata from multiple next-generation sequencing data repositories. *Bioinformatics* **36**, 3607–3609 (2020).
171. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).

172. Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A. & Sun, F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* **5**, 69 (2017).
173. Tithi, S. S., Aylward, F. O., Jensen, R. V. & Zhang, L. FastViromeExplorer: a pipeline for virus and phage identification and abundance profiling in metagenomics data. *PeerJ* **2018**, e4227 (2018).
174. Roux, S., Tournayre, J., Mahul, A., Debroas, D. & Enault, F. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinforma.* **15**, 76 (2014).
175. Jurtz, V. I., Villarroel, J., Lund, O., Voldby Larsen, M. & Nielsen, M. MetaPhinder — identifying bacteriophage sequences in metagenomic data sets. *PLoS ONE* **11**, e0163111 (2016).
176. Rampelli, S. et al. ViromeScan: a new tool for metagenomic viral community profiling. *BMC Genomics* **17**, 165 (2016).
177. Bolduc, B. et al. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect archaea and bacteria. *PeerJ* **2017**, e3243 (2017).
178. Hatcher, E. L. et al. Virus variation resource-improved response to emergent viral outbreaks. *Nucleic Acids Res.* **45**, D482–D490 (2017).
179. Bateman, A. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
180. 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).
181. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
182. Skewes-Cox, P., Sharpton, T. J., Pollard, K. S. & DeRisi, J. L. Profile hidden Markov models for the detection of viruses within metagenomic sequence data. *PLoS ONE* **9**, e105067 (2014).
183. Clarke, E. L. et al. Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. *Microbiome* **7**, 46 (2019).
184. Tisza, M. J. et al. Discovery of several thousand highly diverse circular DNA viruses. *eLife* **9**, e51971 (2020).
185. Zhao, G. et al. VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. *Virology* **503**, 21–30 (2017).
186. McNair, K., Bailey, B. A. & Edwards, R. A. PHACTS, a computational approach to classifying the lifestyle of phages. *Bioinformatics* **28**, 614–618 (2012).
187. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).

Acknowledgements
The authors thank members of the Bushman laboratory for help and suggestions, and L. Zimmerman for artwork designs. This work was supported by the National Institutes of Health (NIH) (grants R61-HL137063, R01-HL113252), the Penn Center for AIDS Research (P30 AI 045008), the PennCHOP Microbiome Program and a Tobacco Formula grant under the Commonwealth Universal Research Enhancement (CURE) programme (grant number SAP # 4100068710), and the Crohn's and Colitis Foundation.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Microbiology thanks F. Maggi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.