

Emerging view of the human virome

KRISTINE M. WYLIE, GEORGE M. WEINSTOCK, and GREGORY A. STORCH

ST. LOUIS, MO

The human virome is the collection of all viruses that are found in or on humans, including both eukaryotic and prokaryotic viruses. Eukaryotic viruses clearly have important effects on human health, ranging from mild, self-limited acute or chronic infections to those with serious or fatal consequences. Prokaryotic viruses can also influence human health by affecting bacterial community structure and function. Therefore, definition of the virome is an important step toward understanding how microbes affect human health and disease. We review progress in virome analysis, which has been driven by advances in high-throughput, deep sequencing technology. Highlights from these studies include the association of viruses with clinical phenotypes and description of novel viruses that may be important pathogens. Together these studies indicate that analysis of the human virome is critical as we aim to understand how microbial communities influence human health and disease. Descriptions of the human virome will stimulate future work to understand how the virome affects long-term human health, immunity, and response to coinfections. Analysis of the virome ultimately may affect the treatment of patients with a variety of clinical syndromes. (*Translational Research* 2012;160:283–290)

Abbreviations: CF = cystic fibrosis; HMP = Human Microbiome Project; RTG = Real Time Genomics

The viral component of the human microbiome is referred to as the “human virome.” The human virome (also referred to as the “viral metagenome”) is the collection of all viruses that are found in or on humans, including viruses causing acute, persistent, or latent infection, and viruses integrated into the human genome, such as endogenous retroviruses. The

human virome includes both eukaryotic and prokaryotic viruses (bacteriophages). Eukaryotic viruses clearly have important effects on human health. Viral infections of humans include acute, self-limited infections; fulminant, uncontrolled acute infections; and chronic infections that may be asymptomatic or associated with serious, even fatal diseases, such as acquired immunodeficiency syndrome.¹ Furthermore, many diseases of unknown cause are thought to be of viral origin.² Human endogenous retroviruses comprise greater than 8% of the human genome.³ They are transcribed ubiquitously in normal tissues.⁴ There has been preliminary evidence of their association with diseases, including amyotrophic lateral sclerosis, multiple sclerosis, and rheumatoid arthritis,^{5–7} however, the association has not been shown to be causal. Bacteriophages may also affect human health because they can influence bacterial population structure or virulence.⁸ Advances in high-throughput, deep sequencing technology make it possible to characterize virome richness and stability, gene functions, and association with disease

From The Genome Institute; Department of Pediatrics, Washington University School of Medicine, St. Louis, Mo.

All authors have read the policy on disclosure of potential conflicts of interest, and the authors have no financial or personal conflicts of interest to disclose.

Submitted for publication December 8, 2011; revision submitted March 28, 2012; accepted for publication March 29, 2012.

Reprint requests: Kristine M. Wylie, Campus Box 8501, The Genome Institute, Washington University School of Medicine, 4444 Forest Park Avenue, St. Louis, MO.

1931-5244

© 2012 Mosby, Inc. Open access under [CC BY-NC-ND license](#).

doi:10.1016/j.trsl.2012.03.006

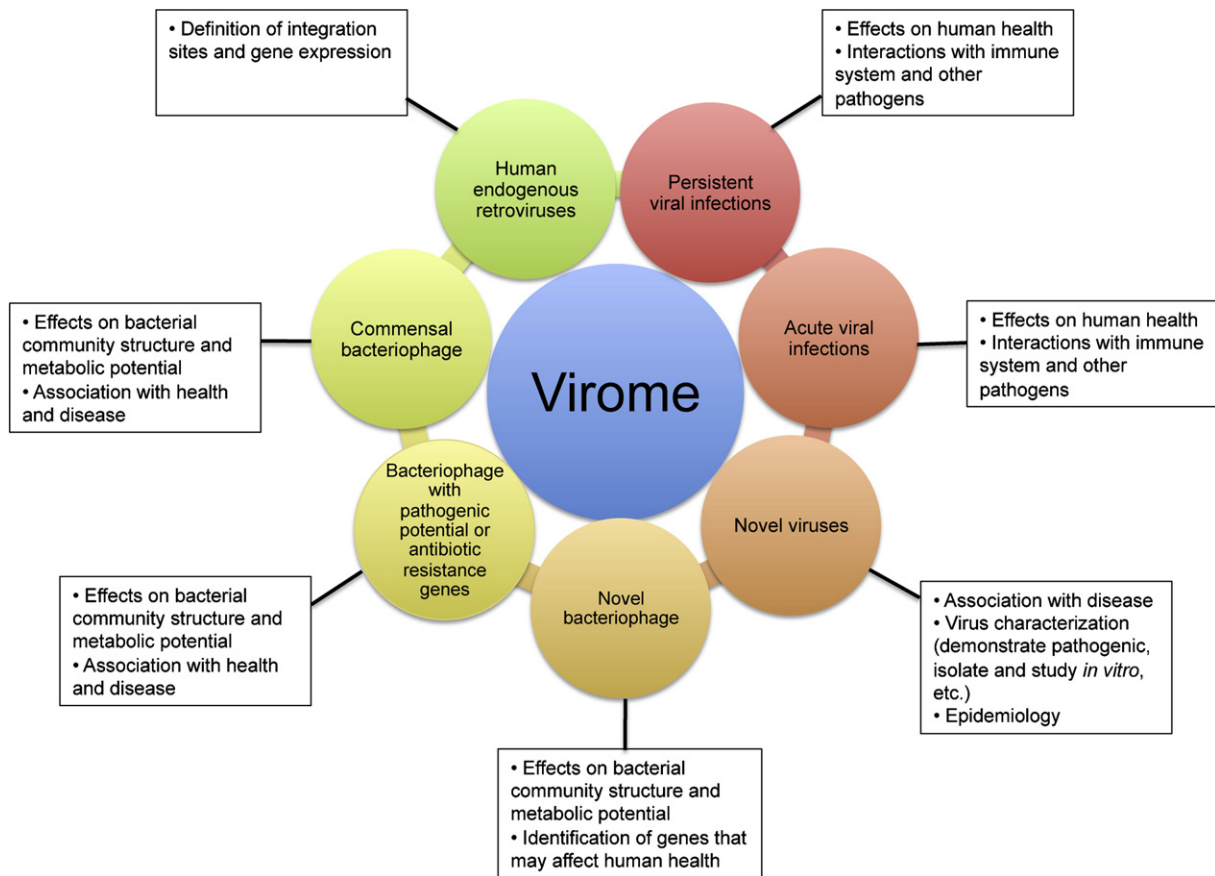


Fig 1. Components of the human virome. Circles represent components of the virome that can be characterized by metagenomic sequencing. These categories are not mutually exclusive. After viruses are identified, their effects on human health and the microbial community must be determined. Examples of questions and implications that are associated with each component of the virome are listed in the boxes.

phenotypes.⁹ Thus, we are poised to begin to understand the richness of the virome and the role viruses play within complex microbial communities (Fig 1).

CHALLENGES FOR VIROME STUDIES

The study of the virome is challenging for several reasons. First, viruses do not contain a conserved genomic region that can be used to identify the viruses in a microbial community, such as the 16S rRNA gene that is used to classify bacteria. Instead, the entire viral community must be sampled and viral genomic sequences compared with known viral reference sequences. The success of this process is currently limited by the fact that many viruses have not yet been characterized and are not included in viral reference databases.¹⁰ Furthermore, viral sequences with poor homology to known viruses may be difficult to classify.

The second challenge in studying the virome is that viral genomic material can be a small proportion of the total nucleic acid in microbial communities because

of the small genome sizes of most viruses and their low-level presence in some cases. This is particularly true for eukaryotic viruses producing persistent asymptomatic infection that may have as yet unappreciated effects on long-term human health.¹¹ Polymerase chain reaction and culture are tools that can be used to characterize the virome. However, the use of these approaches requires up-front decisions about which viruses to look for, thus providing an informative but more limited view of the scope of the virome. Viral nucleic acids can be enriched using hybridization techniques such as microarray or capture,¹²⁻¹⁹ and bound nucleic acids can subsequently be sequenced to provide additional information about the viral genomes. Some novel viruses can be detected by these methods if there is sufficient sequence homology to bind the viral probes.²⁰⁻²³ Enrichment of viral particles via filtration and gradient centrifugation²⁴ can enhance the viral signal. However, enrichment techniques can bias against certain types of viruses, and intracellular and low-abundance viruses

can be lost during the enrichment process.²⁴ High-throughput, deep sequencing technology is revolutionary, because it provides an unbiased approach that can detect even rare components of a microbial community.

Nucleotide sequencing delivers great power for detecting known and novel viruses in clinical samples. Less than 10 years ago, the ABI 3730 capillary sequencer (Applied Biosystems, Foster City, CA) was the state-of-the-art platform for high-throughput sequencing, simultaneously generating sequences from 96 clones on a single run. The lengths of sequences generated on this platform are typically 500 to 800 bases. This relatively long length can be advantageous for discovering novel microbes with remote homologies to reference sequences. However, ABI 3730 sequencing requires that the novel microbe be abundant in the original sample or cloned, because the cost per read limits the number of sequences that can be generated in an experiment. Sequences generated on the ABI 3730 were used for the initial sequence-based characterizations of nonviral microbial communities and for early studies in which novel viral pathogens were detected (discussed below).

In the decade since capillary sequencing was used for the Human Genome Project, technology has increased the yield of sequence that can be generated per day from a single instrument by >30,000-fold while reducing cost by approximately 7000-fold. With the advent of massively parallel sequencing platforms, such as the Roche 454 pyrosequencer (454 Life Sciences, Branford, CT), sequencing capacity grew to approximately 1 million sequences per run, each 250 to 500 bases in length, resulting in a total sequence throughput of up to 500 million bases per run. By introducing sequence “barcodes” during sample amplification, multiple samples can be pooled within a single run, allowing generation of tens to hundreds of thousands of sequences per sample. This massively parallel sequencing allows a more thorough assessment of microbial communities that includes the description of lower abundance microbes. Indeed, analysis of stool samples on the Roche 454 platform revealed a greater number of viruses compared with the ABI 3730.²⁵ Many novel viruses were discovered using the Roche platform (discussed below). The Illumina Genome Analyzer (Illumina Inc, San Diego, CA) generates up to 640 million sequences per run, and the Illumina HiSeq 2000 can generate up to 6 billion paired-end sequences per run. On each of these platforms, multiple pooled, barcoded samples can be included on each run. Illumina sequences are shorter than those generated by Roche 454 pyrosequencing: In early experiments, they were less than 50 bases in length but now are routinely 100 bases. Although the read length is short, sequences can be generated from

both ends of a DNA fragment to yield “paired-end” reads, allowing 200 bases to be sequenced from the same DNA fragment. Illumina technology provides the sensitivity needed to detect rare virus sequences, with sensitivity comparable to that of quantitative reverse transcriptase polymerase chain reaction in some studies.²⁶ The short lengths seem to be sufficient for detecting novel viruses within a sample of a microbial community.²⁷ Assembly of Illumina sequences can also be used to achieve longer contiguous sequences,²⁷ and assembly programs such as PRICE have been developed to extend a fragment of sequence from a novel organism iteratively using paired-end Illumina data (DeRisi, unpublished, available at: <http://derisilab.ucsf.edu/software/price/index.html>). Trends toward increasing numbers of sequences per run and decreased cost per base are likely to continue. New sequencing platforms, including the Illumina MiSeq and the Life Technologies (Grand Island, NY) Ion Torrent Personal Genome Machine Sequencer, are being developed to generate large amounts of sequence data with a rapid turnaround time.

Rapid, accurate analysis of sequence data is critical for research, with more stringent requirements anticipated as clinical applications for virome analysis are developed. Identification of viral sequences is generally achieved by comparison of microbial sequences with reference genomes. Use of programs such as BLAST and BLASTX²⁸ is the traditional method for doing this; these programs work well for relatively small data sets generated by the ABI 3730 and Roche 454 pyrosequencer or for longer contiguous sequences assembled from shorter Illumina reads. However, analyzing millions to billions of Illumina Genome Analyzer sequences requires faster aligners. Many short-read sequence alignment tools are fast but have low tolerance for sequence mismatches; however, virus sequences may differ significantly from the reference genome sequences, so allowing mismatches in the alignments is critical.

Martin and colleagues²⁹ provide a thorough comparison of nucleotide alignment tools for short sequences. CLC bio (www.clcbio.com) and Real Time Genomics (RTG) (www.realtimengenomics.com) software were chosen from the tools evaluated, and they were used extensively to carry out nucleotide alignments of the terabases of Illumina data generated in the Human Microbiome Project (HMP); MBLASTX from Multi Core Ware (www.multicorewareinc.com) and RTG mapx software were used for HMP translated sequence alignments (HMP Consortium, manuscript in revision, 2012). These programs provide 100- to 1000-fold increases in alignment speed over BLAST and BLASTX while maintaining similar sensitivities (MBLASTX,

Mitreva et al, manuscript in revision, 2012) (RTG, Mitreva et al, manuscript in preparation, 2012). Although identification of virus sequences based on sequence homology to known viruses is straightforward in concept, one must be cautious in interpreting the data. Low-complexity sequence and sequences with homology between virus and host can cause false-positive viral identifications. Likewise, false-positive identifications can occur when a sequence does not have close homology to a sequence in the reference database; some general functions are conserved among eukaryotes, bacteria, and DNA viruses, which can result in a weak alignment of translated sequence. Further analysis of virome diversity and complexity can be achieved using software packages, such as GAAS,³⁰ Metavir,³¹ and PHACCS.³² Expertise in the computational challenges of virome analysis will be needed as virome studies become more widespread and move toward clinical applications.

SEQUENCE-BASED VIROME ANALYSIS

Some of the first virome analyses were carried out on environmental samples, particularly those from ocean water.^{33,34} In a study by Breitbart et al,³³ viral DNA was isolated from surface seawater collected in La Jolla and San Diego, California, and approximately 1000 sequences were generated from each sample. Chao1 estimates and rank abundance curves predicted that hundreds to thousands of viral genotypes were present in the viral communities. Significant alignments were identified to all major families of dsDNA tailed phages. In addition, 65% of the sequences were unclassified, pointing to the existence of vast genomic diversity in the oceanic ecosystem, including many novel viruses. Angly et al³⁴ expanded the virome analysis to 4 distinct oceanic regions (Sargasso Sea, Gulf of Mexico, seawater off the coast of British Columbia, and the Arctic ocean) and analyzed samples collected at different time points, locations, and depths. More than 1.7 million sequences were generated using the Roche 454 platform. These sequences were relatively short, with an average length of 102 bases. Oceanic environments contained distinct phage groups that reflected the composition of the bacterial community in that niche, as well as some phages that were common to all or some environments. The diversity and richness of phage populations were different in the 4 environments described. These data suggest that phage communities in different ecologic niches will differ with respect to the environment in which they are found, in part reflecting the resident bacterial population and its functions. This work also suggests that the study of the viral populations in a variety of human body habitats will reveal

an unappreciated diversity of common and specialized viruses.

Early sequence-based analyses of the virome in samples from humans focused on bacteriophage populations. Bacteriophages influence their host bacteria and contribute genes that affect the structure and functions of microbial communities.^{35,36} Therefore, bacteriophages may be both important effectors and indicators of human health and disease. In the first characterization of a bacteriophage community in a human stool sample, shotgun sequencing of 532 cloned viral DNA fragments from the stool of a healthy adult revealed that the majority of phage sequences were novel.³⁷ The data suggested rich diversity of bacteriophage sequences, with approximately 2 to 5 times the number of bacteriophage genotypes as predicted bacterial genera in a stool community (~1200–2000 genotypes predicted).³⁷ In contrast, a simple but dynamic bacteriophage community (~8 genotypes predicted) was observed by sequencing 477 viral DNA clones from feces of a 1-week-old infant.³⁸ These studies suggest that the diversity of bacteriophages in the gut expands as the bacterial community is established,³⁸ but a larger group of adults and infants will need to be sampled and compared to validate this conclusion. In fact, more recent studies that include samples from more individuals and use deeper sequencing indicate that the richness of bacteriophage populations in stool communities varies greatly among adults. In one study, Reyes et al³⁹ found ~10 to 984 genotypes per sample from 12 individuals. In another study, Minot et al⁴⁰ found ~19 to 785 genotypes per sample from 16 individuals. Thus, although important changes in the virome may occur as the infant gut matures, it is likely that the changes are more complex than simply increased diversity.

The insights into the human virome (particularly the bacteriophage component) provided by studies by Reyes et al³⁹ and Minot et al⁴⁰ were made possible in large part because of newer sequencing technologies, especially the Roche 454 pyrosequencing platform. Consistent with the earlier studies, most viral sequences obtained were novel. Both studies showed relative stability of the virome over time (days to years), although changes in diet that affected the bacterial communities also correlated with changes in the viral communities.⁴⁰ The depth of sequencing enabled the assembly of longer contiguous sequences that were used to identify remote homologies and open reading frames for functional analysis. Of importance, the studies by Reyes et al and Minot et al show that bacteriophages encode antibiotic resistance genes⁴⁰ and other genes associated with bacterial metabolic pathways.^{39,40} Also, like bacterial plasmids, bacteriophages serve as reservoirs for mobile genetic elements in bacteria. In turn, this suggests that

bacteriophages may affect human health by contributing to or changing the metabolic capabilities of the resident bacterial community.

The perturbation of a microbial environment by a disease, such as cystic fibrosis (CF), can cause changes in the microbiome. Willner et al⁴¹ studied the viral metagenome of the respiratory tract by analyzing sputum samples from 5 patients with CF and 5 controls without CF.⁴¹ The study describes bacteriophage communities in healthy people that were unique to each individual and were thought to reflect a random, transient sampling of the external environment. However, bacteriophage communities from individuals with CF were similar to each other, presumably driven by effects of their airway pathology. The spouse of a CF patient and a control with asthma, neither of whom had CF, shared the distinct sets of viral taxa and predicted host range found in the individuals with CF. These data lead to 2 important inferences. The first is that environment can have a strong influence on an individual's microbiome, including the virome. In this study, the presence of shared organisms between spouses was striking, indicating a shared external environment. The microbial community was thought to be transient in the spouse without CF but more established in the patient with CF, in whom clearance of microbes is impaired. The second inference is that similar microbial communities may be established in response to distinct health conditions, such as CF and asthma, both of which may cause impaired clearance of microbes from the airways. Together, these data suggest that in addition to the components of the virome, the dynamics of the viral community may be important for distinguishing the effects of the virome in different microenvironments.

The studies discussed thus far evaluated DNA viruses, but many important RNA viruses that infect eukaryotic cells also are found in the gastrointestinal and respiratory tracts. Eukaryotic viruses are found less frequently than bacteriophages in many microbial communities, and indeed the stool and sputum samples evaluated contained only a few sequences with homology to eukaryotic DNA viruses.³⁹⁻⁴¹ It is likely that more eukaryotic viruses would be found by inclusion of RNA in the analysis (particularly in the respiratory tract). A study that evaluated RNA viruses in stool samples from 2 healthy individuals found a diverse array of viruses.⁴² Although some human viruses (picobirnaviruses) were detected in this study, most of the RNA viruses detected were plant viruses, which were most likely found in the stool as a by-product of diet.

Raw sewage and reclaimed water provide source material for virus discovery and the evaluation of emerging pathogens.⁴³⁻⁴⁵ DNA and RNA virus sequences from raw sewage collected at several sites⁴³ revealed a viral

community that was dominated by bacteriophages and the subset of eukaryotic viruses that were predominantly from plants. Seventeen known human viruses were detected. Strikingly, novel viruses belonging to 51 virus families were also detected. These data indicate that environmental samples that contain specimens from a large number of individuals can provide valuable information concerning viruses present in the population, including novel agents in addition to known human pathogens.

VIROME AND DISEASE: THE DISCOVERY OF NOVEL EUKARYOTIC VIRUSES

Overall, eukaryotic viruses are minor components of a microbial community, although their effects are often readily observed. Titers of eukaryotic viruses are generally higher in samples from symptomatic versus asymptomatic individuals. Thus, some of the viral metagenomic studies of the human gastrointestinal tract evaluated stool from patients with diarrhea⁴⁶ and non-polio acute flaccid paralysis.²⁵ The samples evaluated (from 12 and 35 patients, respectively) contained a variety of DNA and RNA viruses, including human enteroviruses, adenoviruses, caliciviruses, and parvoviruses. The eukaryotic viral metagenomes were distinct in each subject. Viral sequences accounted for the majority of sequences that were present in some subjects. The use of the Roche 454 pyrosequencing platform, which generated more sequences per sample than the ABI 3730 platform, revealed a greater richness in the eukaryotic viral metagenome.²⁵ This indicates that depth of sampling is an important factor for comprehensive viral metagenomic analysis and for discovering novel eukaryotic viruses. In addition to the detection of known viruses, each of these studies identified novel viruses associated with diarrhea, including an astrovirus,⁴⁶ a cosavirus, and a bocavirus,²⁵ among others. Novel viruses identified by these viral metagenomic studies must be subject to extensive further study to determine whether they are causally associated with human disease.⁴⁷

The identification of novel viruses is an exciting part of the characterization of the virome. Most of the viral sequences detected in deep sequencing experiments are uncharacterized (described above), indicating the presence of great viral diversity to be discovered. These undiscovered viruses may affect human health, either acutely or through chronic infection.¹¹ Indeed, many conditions, including fever, diarrhea, and respiratory illness, may be caused by unknown or undiagnosed pathogens that are suspected to be viral. In recent years, many novel eukaryotic viruses have been discovered or characterized using sequencing, including viruses in the following groups: arenaviruses,⁴⁸ astroviruses,^{46,49,50}

rhinoviruses,⁵¹ nodoviruses,²⁵ coronaviruses,²⁰ polyomaviruses,^{52,53} bocaviruses,⁵⁴ enteroviruses,⁵⁵ and klasseviruses,⁵⁶ among others. Many more viruses undoubtedly remain to be discovered, and further characterization of viral strains and subtypes is an important goal.⁵⁷

Discoveries about the presence and dynamics of known viruses in the virome may also affect the way we view their impact on human health. For instance, viruses that integrate into the human genome have been associated with cancer (eg, human papillomavirus 16, Epstein–Barr virus, and the more recently discovered Merkel cell polyomavirus). As we characterize the human virome, distinguishing episomal from integrated viruses is an important goal that may relate to the understanding of disease. In addition, virome analysis may identify known viruses in unexpected tissues, which could suggest novel mechanisms of disease.

THE FUTURE: CLINICAL APPLICATIONS OF HUMAN VIROME STUDIES

The most immediate applications of virome studies relate to the discovery of new viral pathogens (see above) or viruses with previously unappreciated tropisms.^{58,59} Ongoing viral metagenomic analyses will undoubtedly reveal the presence of additional novel viruses. Significant evidence must be accrued to relate novel viruses to disease phenotypes. As evidence associating novel viruses with disease phenotypes accumulates, these new viruses will be considered as potential causes for disease. For instance, since their discovery in 2005,⁵⁴ bocaviruses have been associated with respiratory illness and diarrhea,⁶⁰ however, their roles as pathogens have not yet been formally established. Detailed studies will be required to establish causal relationships between viruses and disease.

An intriguing question is whether viral metagenomic analysis can be applied as a clinical diagnostic method. The concept is appealing because a sequencing-based approach could dramatically increase the range of viruses detected in clinical samples compared with existing diagnostic methods. In some recent studies, sequence-based analysis of viral communities has had sensitivity comparable to virus-specific polymerase chain reaction.²⁶ Alternative approaches would be to enrich for viral nucleic acids by carrying out hybridization or alternatively to remove human nucleic acid before sequencing.^{12–14} Important methodological questions that need to be addressed include which samples should be selected for analysis, what sample preparation method should be used, and which sequencing platform should be used. In addition, extensive work remains to be done by laboratorians and clinicians to understand the

clinical significance of the data generated. Finally, significant practical barriers remain to be surmounted, including decreasing the time required for sample-to-result analysis and decreasing cost. Although further technologic progress in both sequencing and information processing will be required to meet these goals, the pace of recent advances suggests that this may occur in the relatively near future.

CONCLUSIONS

We envision that in some patients who are diagnostic mysteries, rapid, unbiased sequence analysis of the viral metagenome in several samples from the patient will be used to generate a list of medically relevant viruses and genes that are detected, which can be further evaluated and confirmed using virus-specific assays. The viral metagenomic data will then be considered along with clinical data to determine whether (a) the virus or viruses can have a causal relationship to the patient's illness or (b) genes encoded by the virus may affect a planned treatment (antibiotic or antiviral resistance). In the future, as we begin to understand how the virome affects long-term human health, immunity, and response to coinfections or treatments, analysis of the virome may become highly informative for patient management.

REFERENCES

1. Fields BN, Knipe DM, Howley PM. *Fields' Virology*. 5th ed. Philadelphia: Lippincott-Williams & Wilkins, 2007.
2. Foxman EF, Iwasaki A. Genome-virome interactions: examining the role of common viral infections in complex disease. *Nat Rev Microbiol* 2011;9:254–64.
3. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
4. Seifarth W, Frank O, Zeilfelder U, et al. Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *J Virol* 2005;79:341–52.
5. Douville R, Liu J, Rothstein J, Nath A. Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. *Ann Neurol* 2011;69:141–51.
6. Balada E, Vilardell-Tarres M, Ordi-Ros J. Implication of human endogenous retroviruses in the development of autoimmune diseases. *Int Rev Immunol* 2010;29:351–70.
7. Christensen T. HERVs in neuropathogenesis. *J Neuroimmune Pharmacol* 2010;5:326–35.
8. Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 1996;272:1910–4.
9. Haynes M, Rohwer F. The human virome. *Metagenomics of the human body*. 2011;10.1007/978-1-4419-7089-3_4:63–77.
10. Woolhouse ME, Howey R, Gaunt E, Reilly L, Chase-Topping M, Savill N. Temporal trends in the discovery of human viruses. *Proc Biol Sci* 2008;275:2111–5.
11. Virgin HW, Wherry EJ, Ahmed R. Redefining chronic viral infection. *Cell* 2009;138:30–50.

12. Wang D, Coscoy L, Zylberberg M, et al. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A* 2002;99:15687–92.
13. Depledge DP, Palser AL, Watson SJ, et al. Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One* 2011;6:e27805.
14. Duncavage EJ, Magrini V, Becker N, et al. Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. *J Mol Diagn* 2011;13:325–33.
15. Chen EC, Miller SA, DeRisi JL, Chiu CY. Using a pan-viral microarray assay (Virochip) to screen clinical samples for viral pathogens. *J Vis Exp* 2011;(50). pii: 2536. doi: 10.3791/2536.
16. Korimbocus J, Scaramozzino N, Lacroix B, Crance JM, Garin D, Vernet G. DNA probe array for the simultaneous identification of herpesviruses, enteroviruses, and flaviviruses. *J Clin Microbiol* 2005;43:3779–87.
17. Lin B, Wang Z, Vora GJ, et al. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. *Genome Res* 2006;16:527–35.
18. Wong CW, Heng CL, Wan Yee L, et al. Optimization and clinical validation of a pathogen detection microarray. *Genome Biol* 2007;8:R93.
19. Palacios G, Quan PL, Jabado OJ, et al. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg Infect Dis* 2007;13:73–81.
20. Wang D, Urisman A, Liu YT, et al. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 2003;1:E2.
21. Chen EC, Yagi S, Kelly KR, et al. Cross-species transmission of a novel adenovirus associated with a fulminant pneumonia outbreak in a new world monkey colony. *PLoS Pathog* 2011;7:e1002155.
22. Chiu CY, Greninger AL, Kanada K, et al. Identification of cardioviruses related to Theiler's murine encephalomyelitis virus in human infections. *Proc Natl Acad Sci U S A* 2008;105:14124–9.
23. Mihindukulasuriya KA, Wu G, St Leger J, Nordhausen RW, Wang D. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J Virol* 2008;82:5084–8.
24. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. *Nat Protoc* 2009;4:470–83.
25. Victoria JG, Kapoor A, Li L, et al. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol* 2009;83:4642–51.
26. Cheval J, Sauvage V, Frangeul L, et al. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J Clin Microbiol* 2011;49:3268–75.
27. Wylie KM, Mihindukulasuriya KA, Sodergren E, Weinstock GM, Storch GA. Sequence analysis of the human virome in febrile and afebrile children. *PLoS One* 2012;7:e27735.
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.
29. Martin J, Sykes S, Young S, et al. Optimizing read mapping to reference genomes to determine composition and species prevalence in microbial communities. *PLoS One* 2012;7:e36427.
30. Angly FE, Willner D, Prieto-Davo A, et al. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS Comput Biol* 2009;5:e1000593.
31. Roux S, Faubladier M, Mahul A, et al. Metavir: a web server dedicated to virome analysis. *Bioinformatics* 2011;27:3074–5.
32. Angly F, Rodriguez-Brito B, Bangor D, et al. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* 2005;6:41.
33. Breitbart M, Salamon P, Andresen B, et al. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* 2002;99:14250–5.
34. Angly FE, Felts B, Breitbart M, et al. The marine viromes of four oceanic regions. *PLoS Biol* 2006;4:e368.
35. Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brussow H. Phage-host interaction: an ecological perspective. *J Bacteriol* 2004;186:3677–86.
36. Brussow H, Canchaya C, Hardt WD. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 2004;68:560–602.
37. Breitbart M, Hewson I, Felts B, et al. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* 2003;185:6220–3.
38. Breitbart M, Haynes M, Kelley S, et al. Viral diversity and dynamics in an infant gut. *Res Microbiol* 2008;159:367–73.
39. Reyes A, Haynes M, Hanson N, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 2010;466:334–8.
40. Minot S, Sinha R, Chen J, et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 2011;21:1616–25.
41. Willner D, Furlan M, Haynes M, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One* 2009;4:e7370.
42. Zhang T, Breitbart M, Lee WH, et al. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol* 2006;4:e3.
43. Cantalupo PG, Calgua B, Zhao G, et al. Raw sewage harbors diverse viral populations. *MBio* 2011;2.
44. Rosario K, Duffy S, Breitbart M. Diverse circovirus-like genome architectures revealed by environmental metagenomics. *J Gen Virol* 2009;90:2418–24.
45. Symonds EM, Griffin DW, Breitbart M. Eukaryotic viruses in wastewater samples from the United States. *Appl Environ Microbiol* 2009;75:1402–9.
46. Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog* 2008;4:e1000011.
47. Finkbeiner SR, Le BM, Holtz LR, Storch GA, Wang D. Detection of newly described astrovirus MLB1 in stool samples from children. *Emerg Infect Dis* 2009;15:441–4.
48. Palacios G, Druce J, Du L, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med* 2008;358:991–8.
49. Finkbeiner SR, Li Y, Ruone S, et al. Identification of a novel astrovirus (astrovirus VA1) associated with an outbreak of acute gastroenteritis. *J Virol* 2009;83:10836–9.
50. Kapoor A, Li L, Victoria J, et al. Multiple novel astrovirus species in human stool. *J Gen Virol* 2009;90:2965–72.
51. McErlean P, Shackelton LA, Lambert SB, Nissen MD, Sloots TP, Mackay IM. Characterisation of a newly identified human rhinovirus, HRV-QPM, discovered in infants with bronchiolitis. *J Clin Virol* 2007;39:67–75.
52. Allander T, Andreasson K, Gupta S, et al. Identification of a third human polyomavirus. *J Virol* 2007;81:4130–6.
53. Gaynor AM, Nissen MD, Whiley DM, et al. Identification of a novel polyomavirus from patients with acute respiratory tract infections. *PLoS Pathog* 2007;3:e64.
54. Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci U S A* 2005;102:12891–6.

55. Yozwiak NL, Skewes-Cox P, Gordon A, et al. Human enterovirus 109: a novel interspecies recombinant enterovirus isolated from a case of acute pediatric respiratory illness in Nicaragua. *J Virol* 2010;84:9047–58.
56. Greninger AL, Runckel C, Chiu CY, et al. The complete genome of klassevirus - a novel picornavirus in pediatric stool. *Virology* 2009;6:82.
57. Peterson J, Garges S, Giovanni M, et al. The NIH Human Microbiome Project. *Genome Res* 2009;19:2317–23.
58. Holtz LR, Wylie KM, Sodergren E, et al. Astrovirus MLB2 viremia in febrile child. *Emerg Infect Dis* 2011;17:2050–2.
59. Wunderli W, Meerbach A, Guengoer T, et al. Astrovirus infection in hospitalized infants with severe combined immunodeficiency after allogeneic hematopoietic stem cell transplantation. *PLoS One* 2011;6:e27483.
60. Chow BD, Esper FP. The human bocaviruses: a review and discussion of their role in infection. *Clin Lab Med* 2009;29:695–713.