



Published in final edited form as:

Curr Opin Virol. 2011 October 1; 1(4): 282–288. doi:10.1016/j.coviro.2011.07.006.

From orphan virus to pathogen: the path to the clinical lab

Linlin Li^{1,2} and Eric Delwart^{1,2,*}

¹Blood Systems Research Institute, San Francisco, CA, USA

²Department of Laboratory Medicine, University of California, San Francisco, CA, USA

Abstract

Viral metagenomics has recently yielded numerous previously uncharacterized viral genomes from human and animal samples. We review some of the metagenomics tools and strategies to determine which orphan viruses are likely pathogens. Disease association studies compare viral prevalence in patients with unexplained symptoms versus healthy individuals but require these case and control groups to be closely matched epidemiologically. The development of an antibody response in convalescent serum can temporarily link symptoms with a recent infection. Neutralizing antibody detection require often difficult cell culture virus amplification. Antibody binding assays require proper antigen synthesis and positive control sera to set assay thresholds. High levels of viral genetic diversity within orphan viral groups, frequent co-infections, low or rare pathogenicity, and chronic virus shedding, can all complicate disease association studies. The limited availability of matched cases and controls sample sets from different age groups and geographic origins is a major block for estimating the pathogenic potential of recently characterized orphan viruses. Current limitations on the practical use of deep sequencing for viral diagnostics are listed.

Keywords

orphan virus; viral metagenomics; disease association; virus discovery

Introduction

Viral metagenomics directly characterizes the genetic material of viral communities bypassing the need for prior virus-specific *in vitro* or *in vivo* amplification [1–3]. Viral metagenomics was initially used to analyze environmental samples using Sanger shotgun sequencing [4], and has rapidly expanded to include samples such as human and animal feces [5–11], blood [12,13], tissues [14–17] and respiratory secretions [18–20] often using next generation sequencing. Viral metagenomics “deep sequencing” has been largely focused on viral discovery to identify new pathogens, sequencing of viral variants of known species to better understand their evolution [21], and unbiased survey of viral communities [22–25] without the diversity-reducing effect of prior amplification in cell culture. Viral metagenomics can also be used to detect low titer virus and characterize full viral genomes

© 2011 Elsevier B.V. All rights reserved.

*Corresponding author: delwarte@medicine.ucsf.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

circumventing the need for molecular methods requiring stringent nucleotide hybridization such as PCR and microarrays [26].

Sequencing

Due to its earlier introduction and longer read length (facilitating the recognition of highly divergent viral sequences)[27] the Roche 454 system has been the most popular high-throughput sequencing tool for viral discovery. The less expensive (per base pair) shorter sequence reads from the Illumina genome analyzer platform are gradually increasing in popularity [28,29]. Sequencing technologies delivering data in hours instead of days, including that from Ion Torrent and Pacific Biosciences, have been recently introduced but yield fewer reads. A high rate of sequencing errors is inherent to many of these high throughput technologies. Frame-shifting insertion/deletion (indel) mutations disrupt viral open reading frames and thus interfere with protein similarity searches that can impact the identification of highly divergent viruses. Indels have less impact on the identification of already known viral pathogens possible using nucleotide similarity searches. When sequencing high titer viruses sequencing errors are corrected during assembly of the consensus sequence genome from the numerous overlapping sequences covering each nucleotide positions. Deep sequencing may also be used to detect rare variants such as HIV drug resistance mutants present as a minority within largely wild-type viral quasispecies [30].

Viral particles enrichment

Directly sequencing nucleic acids in biological fluid without prior viral particle enrichment will result in a high background of host and bacterial genetic material including chromosomal, episomal and ribosomal sequences [31]. To reduce such background viruses can be purified using simple filtration methods to exclude larger particles. Nuclease treatments can also be used to digest naked cellular nucleic acids abundant in biological fluids while viral nucleic acids remain protected within the viral capsid [31]. When large sample volume is available, such as in environmental studies, ultra-centrifugation methods can also be used to concentrate and purify viral particles from the expected density bands [32,33].

Random amplifications

Regardless of the sequencing technology used, the viral nucleic acids have to be amplified to generate the large quantities of DNA required for most sequencing platforms. Since many viruses have RNA genomes, their detection first requires the genomic RNA to be reverse transcribed into cDNA. Numerous sequence-independent amplification methods have been successfully used but two basic approaches predominate [3,21,34–41]. Random PCR is based on the degenerate 3' end of PCR primers used to randomly prime DNA synthesis [20,42]. The 3' degeneracy allows such primers to anneal throughout the length of viral RNA or DNA genomes. Following two rounds of extension placing such a primer at both extremities of a viral sequence, multiple rounds of PCR amplification are performed with the same primer but lacking the degenerate 3' end (NNNNNN). This method is analogous to one commonly used for generating fluorescently labeled cDNA for analysis of transcriptome on microarrays [43]. A related technique called sequence-independent single primer amplification (SISPA) involves partial cleavage of double stranded cDNA (generated with a short random primer) using a 4bp recognizing restriction enzyme followed by ligation of a sticky adaptor and PCR using a primer complementary to the adaptor [31,39,44].

Another popular random amplification method involves the use of the highly processive phix29 DNA polymerase with short random primers in a multiple displacement

amplification [45,46]. A possible bias for the amplification of circular DNA viral genome may occur using phix29. Viral RNA can also be amplified after reverse transcription and ligation of the cDNA molecules into long chimeric DNA molecules more appropriate for phix29 DNA polymerase amplification [45].

Sequence analysis and viral genome acquisition

To facilitate the recognition of novel viruses, overlapping sequence reads are first computationally assembled into longer contigs. The less complex the nucleic acid population, the greater the number of reads that will be collapsed into contigs. Both contigs and unassembled singletons are then subjected to searches against public sequence databases using the BLAST suite of tools [47]. Nucleotide similarity can quickly identify sequences that are closely related to those of known viral species. Previously uncharacterized viral species that are highly divergent from those already in public databases may be unrecognizable using nucleotide similarity searches and require the computationally more demanding comparison of their virtual translation products to all known viral proteins in order to detect weaker matches. A limitation of such similarity-based approach is that still uncharacterized viral families with no detectable protein sequence similarity to known viral families in GenBank would go unidentified. Possible approaches to identify such highly divergent viruses include further analysis of the taxonomically unclassifiable contigs often found in viral metagenomic studies. Extension of such contigs by targeted PCR or further metagenomics sequencing may allow the detection of distant homology to a known virus. The presence of the same unclassifiable sequence in different patients might also support claims for the detection of a new viral family but would ultimately require direct evidence of viral replication such as amplification *in vitro* or *in vivo*, and sero-conversion in exposed hosts. Once a single genome of a new viral family is characterized its constituent genera and species will be readily detectable using sequence similarity searches. Characterization of novel viral families from any particular host group such as vertebrates, arthropods, protists, plants, and prokaryotes will therefore facilitate detection of their viral homologues in other hosts.

From genomes to the clinical lab

A significant fraction of common human diseases including acute gastroenteritis, acute respiratory tract infections, hepatitis, and encephalitis remain without infectious etiology. Autoimmune diseases and some cancers may also be induced by still unidentified viral triggers. Viral metagenomics therefore provides a simple tool to identify candidate pathogens for such diseases. Clinical diagnostic laboratories frequently use cell culture and on occasion observe viral cytopathic effects induced by unidentifiable viruses. When viral concentration is high, random amplification and even superficial DNA sequencing (i.e. a few plasmid subclones) can rapidly characterize the viral genomes in cell culture supernatants [48,49]

Since its recent deployment, viral metagenomics has quickly led to the genetic characterization of numerous human and animal viruses. While many of these “new” viruses were discovered in clinical samples from patients with unexplained symptoms, their identification in such context may be coincidental and may simply reflect harmless and very common infections. Evidence of the pathogenicity of such orphan viruses is therefore required before these viruses may be included in diagnostic tests. Diagnosis of infection with a particular viral pathogen against which no treatment exists may prevent unnecessary intervention such as antibiotic treatment, and improve palliative measures and transmission prevention.

Demonstrating pathogenicity

What data are required before an infectious agent can be generally accepted as a pathogen has been previously reviewed in depth [50,51]. Human inoculation with orphan viruses being precluded, indirect evidence of their pathogenicity relies largely on observation of natural infections. The initial question is often whether a newly characterized virus is actually a human virus or is simply ingested or inhaled and passing without consequence through gut or bronchi lumen. The common detection of ingested plant, animal, and insect viral nucleic acids in human and animal stools attests to their capacity to survive through the digestive track [6,7,52–54]. Detection of virus specific human antibodies can be used to show replication in human hosts [55,56]. Virus detection in blood, tissues, or CSF (as opposed to the gut, respiratory secretions, or skin) may also be considered evidence of human replication since it is difficult to conceive how a large number of viral particles could be passively transferred to internal anatomical compartments. Viral replication in human cell culture may also reflect a human tropism but because host restrictions can be bypassed in cell cultures and many viruses can grow in non-host species cell lines *in vitro*, replication cannot be considered definitive evidence.

Disease association studies

Determining whether a human virus is pathogenic often starts by comparing its prevalence in matched disease cases and healthy controls using PCR. Single or nested PCR or quantitative PCR assays can be rapidly designed for sensitive detection and viral load measurement. A major block to performing such case-control studies is acquiring large numbers of samples from both unexplained disease cases and from healthy controls, samples not typically collected in medical care facilities. As an alternative, control samples may be collected from patients with unrelated symptoms. Because viral exposure and susceptibility can vary greatly in different human populations the cases and controls need to be carefully matched epidemiologically especially for age, geographic provenance, and if possible general exposure to viruses (socio-demographics or occupation). Failure to properly match cases with controls may lead to misleading results when comparing groups that have different levels of harmless commensal virus infections. Disease association studies would optimally also be performed with different age groups from different continents. The results of such association studies can be strongly suggestive of a pathogenic role.

Disease outbreaks

Samples from clustered disease outbreaks may also be used to test disease association. If a virus is present in all or most patients within an outbreak but not, or only rarely, in unaffected local controls it may be considered associated with the symptoms, especially if other potential infectious culprits are shown to be absent. Single disease cases may also be highly informative if longitudinal samples are collected showing a rise and decline in the concentration of new virus coinciding with onset and resolution of symptoms and known pathogens are not detected using either a battery of specific tests or viral metagenomics.

Antibody sero-conversion

Seroconversion in convalescent sera is also often used to time an infection to the onset of symptoms [57]. For some infections such as HCV and HIV viral replication and particle release may linger for long periods and even establish life-long chronic infections. Some parvoviruses may even be retained *in vivo* in likely inert but PCR detectable forms for years [58,59]. Demonstrating seroconversion in convalescent sera through IgM or IgG detection with a recent prior antibody negative bleed can temporarily associate an infection with onset of symptoms [57]. Both neutralizing and viral antigen binding antibodies may be used.

Neutralizing antibody tests require prior viral amplification in culture, which may be problematic for some viruses. Antibody-antigen binding assays require generating properly folded and modified viral epitopes and assay calibration typically using known antibody positive sera, reagents that may not be readily available. The absence of such positive and negative control sera may be circumvented by setting a threshold level based on bimodal distribution of antibody reactivity within a population or based on the inflection point of ranked binding assay signals [60,61].

The detection of viral replication or expression in affected tissue(s) such as the liver for hepatitis, and brain or CSF for encephalitis can also provide supporting evidence for a pathogenic role. The strong association of the recently described Merkel cell polyomavirus (MCPV) [62] with Merkel cell carcinoma [63] provides an example of convincing disease association for an otherwise highly prevalent infection [61,64]. The virus is found clonally integrated into the cells of the tumor with viral mutations that truncate a helicase gene, presumably to promote transformed cell survival by preventing viral replication [65].

Complications

Complicating disease association studies are the extent of viral genetic diversity shown by some recently characterized viral groups [66–69]. As an example, a high number of serotype/genotype can be found within each human enterovirus species and each associated with no or a variety of symptoms (http://www.picornaviridae.com/enterovirus/prototypes/hev-b_prototypes.htm) [70]. Other groups of similarly genetically divergent viruses may therefore be as equally diverse phenotypically. When determining the pathogenicity of such new viral groups each serotype/genotype should therefore be considered separately increasing the number of samples required to detect sufficient numbers of infections to measure possible disease association.

Many viruses are common but only cause symptoms in a very small subset of infections. To detect a disease association for a virus frequently detected in healthy controls, it must be found in a still higher fraction of the unexplained matched disease cases. If a virus is responsible for only a small fraction of the unexplained cases it may be difficult for that increase in prevalence to rise above the high background of asymptomatic infections. Two groups of very common human viruses found in plasma, the anellovirus genus and GBV-C in the *Flaviviridae* family, were initially detected in hepatitis patients and at first thought to be associated with that condition [71–73]. Further studies showed anelloviruses to be extremely diverse and nearly universal in human plasma and likely transmitted through close family contact very early after birth [74,75]. GBV-C was also shown to be a very common infection, especially in individuals exposed to blood products [76]. Association of either group of virus with hepatitis was not confirmed [75,77,78]. Given that anelloviruses exhibit an extreme degree of genetic variability it remains possible that a subset are associated with some disease in a situation analogous to those of papillomaviruses, but their ubiquitous detection makes association studies challenging [79].

Co-infections may also aggravate symptoms. If disease induction is greatly increased in the context of other infections then only the total number of infections or particular combinations of viruses may be associated with symptoms. Deciphering such complex interactions will require metagenomics or microarray technologies that query all virus families rather than single virus assays in order to identify all participating viruses.

Ultimately, convincingly demonstrating disease association requires confirmation by different research groups using different sets of patient and control samples. The reduction of symptoms and disease prevalence seen after using specific anti-virals and vaccinations

are direct ways to demonstrate human virus pathogenicity. Since such measures are only developed for highly pathogenic and prevalent viral infections, the ultimate confirmation of their pathogenicity first depends on disease association tests such as described above. For animal viruses their direct inoculation into their host species can greatly facilitate the determination of their pathogenicity.

Viral metagenomics in the clinical lab

The use of viral metagenomics for diagnostic purposes appears promising once significant roadblocks are removed [1,26,80]. Meanwhile deep sequencing can be used to detect viral contamination during the development and manufacture of biological products [81,82]. The current barriers to the routine use of viral metagenomics as a diagnostic tool include its current complexity and slow speed, difficult data interpretation, and access to more streamlined alternative methods such as qPCR, or microarrays [83,84]. The random amplification and DNA library construction protocols are currently complex but likely to become simplified and faster in the near future. The sequencing time itself is also likely to be reduced. Shrinking the time from clinical sample collection to sequence data generation and bioinformatics analysis to a medically useful few hours to a day will enhance the appeal of a sequencing approach for diagnostic purposes. The issue of sensitivity to allow the detection of very few copies of viral genomes will also have to be compared, using spiked samples, to highly sensitive PCR based and microarray methods. Cost will likely remain a major barrier especially when compared to a single qPCR assay in cases where only testing for a single virus is indicated. Further limitations include access to appropriate computational power needed to analyze very large sequence data sets. What level of sequence divergence to a reference viral genome (i.e. BLAST E score) will be tolerable and how many matches to different regions of the genome will be required to confirm the presence of a virus remain to be determined. The persistent problem of DNA contamination and the medical significance of very low level of viral nucleic acid detection will also need to be resolved. If the issues described above can be successfully addressed, deep sequencing does have the potential to replace many virus-specific tests by a single procedure. In the immediate future, microbial sequences microarrays are likely to be simpler, cheaper, and more rapid tools to rapidly screen clinical samples in an unbiased way for all of the rapidly growing number of already genetically characterized viruses. In the interim, deep sequencing is likely to be more popular for research use such as viral discovery, re-sequencing, and testing biological products during early development.

Bullet points

- 1.** Metagenomics is identifying many viruses of unknown pathogenicity.
- 2.** Disease association requires epidemiologically well-matched cases and healthy controls groups.
- 3.** Sero-conversion can temporarily link infection with symptoms.
- 4.** Viral genetic diversity and co-infections can complicate disease association studies.
- 5.** Metagenomics for rapid diagnosis of infectious agents is premature but promising.

Acknowledgments

The project described was supported by Award Numbers R01HL105770 and R01HL083254 from the NHLBI and support from BSRI. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHLBI.

Bibliography

1. Tang P, Chiu C. Metagenomics for the discovery of novel human viruses. *Future Microbiol.* 2010; 5:177–189. [PubMed: 20143943]
2. Delwart EL. Viral metagenomics. *Rev Med Virol.* 2007; 17:115–131. [PubMed: 17295196]
3. Bexfield N, Kellam P. Metagenomics and the molecular identification of novel viruses. *Vet J.* 2010 Review describing different random amplification and sequence similarity based methods for viral discovery.
4. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A.* 2002; 99:14250–14255. [PubMed: 12384570]
5. Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, Rohwer F. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol.* 2003; 185:6220–6223. [PubMed: 14526037]
6. Li L, Victoria JG, Wang C, Jones M, Fellers GM, Kunz TH, Delwart E. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J Virol.* 2010; 84:6955–6965. [PubMed: 20463061]
7. Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, Naeem A, Zaidi S, Delwart E. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol.* 2009; 83:4642–4651. [PubMed: 19211756]
8. Finkbeiner SR, Li Y, Ruone S, Conrardy C, Gregoricus N, Toney D, Virgin HW, Anderson LJ, Vinjé J, Wang D, et al. Identification of a novel astrovirus (astrovirus VA1) associated with an outbreak of acute gastroenteritis. *J Virol.* 2009; 83:10836–10839. [PubMed: 19706703]
9. 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.
10. Cann AJ, Fandrich SE, Heaphy S. Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes. *Virus Genes.* 2005; 30:151–156. [PubMed: 15744573]
11. Kapoor A, Slikas E, Simmonds P, Chieochansin T, Naeem A, Shaikat S, Alam MM, Sharif S, Angez M, Zaidi S, et al. A Newly Identified Bocavirus Species in Human Stool. *J Infect Dis.* 2009; 199:196–200. [PubMed: 19072716]
12. Jones MS, Kapoor A, Lukashov VV, Simmonds P, Hecht F, Delwart E. New DNA viruses identified in patients with acute viral infection syndrome. *J Virol.* 2005; 79:8230–8236. [PubMed: 15956568]
13. Breitbart M, Rohwer F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques.* 2005; 39:729–736. [PubMed: 16312220]
14. Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL. Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathog.* 2008; 4 e1000163.
15. Palacios G, Lovoll M, Tengs T, Hornig M, Hutchison S, Hui J, Kongtorp RT, Savji N, Bussetti AV, Solovyov A, et al. Heart and skeletal muscle inflammation of farmed salmon is associated with infection with a novel reovirus. *PLoS One.* 2010; 5 e11487.
16. Kistler AL, Gancz A, Clubb S, Skewes-Cox P, Fischer K, Sorber K, Chiu CY, Lublin A, Mechani S, Farnoushi Y, et al. Recovery of divergent avian bornaviruses from cases of proventricular dilatation disease: identification of a candidate etiologic agent. *Virol J.* 2008; 5:88. [PubMed: 18671869]
17. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, Khristova ML, Weyer J, Swanepoel R, Egholm M, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog.* 2009; 5 e1000455.

18. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, Tammadoni S, Nosrat B, Conrad D, Rohwer F. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One*. 2009; 4 e7370.
19. Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, Persson MA, Dalianis T, Ramqvist T, Andersson B. Identification of a third human polyomavirus. *J Virol*. 2007; 81:4130–4136. [PubMed: 17287263]
20. 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; 102:12891–12896. [PubMed: 16118271]
21. Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N, Afonso C, Zhang X, Anderson NG, Ghedin E, et al. Viral genome sequencing by random priming methods. *BMC Genomics*. 2008; 9:5. [PubMed: 18179705]
22. Djikeng A, Kuzmickas R, Anderson NG, Spiro DJ. Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One*. 2009; 4 e7264.
23. Culley AI, Lang AS, Suttle CA. Metagenomic analysis of coastal RNA virus communities. *Science*. 2006; 312:1795–1798. [PubMed: 16794078]
24. Culley AI, Lang AS, Suttle CA. High diversity of unknown picorna-like viruses in the sea. *Nature*. 2003; 424:1054–1057. [PubMed: 12944967]
25. Runckel C, Flenniken ML, Engel JC, Ruby JG, Ganem D, Andino R, DeRisi JL. Temporal Analysis of the Honey Bee Microbiome Reveals Four Novel Viruses and Seasonal Prevalence of Known Viruses, Nosema, and Crithidia. *PLoS ONE*. 2011; 6 e20656–e20656.
26. Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, Kim E, Pillai DR, Guyard C, Mazzulli T, et al. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One*. 2010; 5 e13381. First report of the diagnostic sensitivity of deep sequencing approaching that of PCR for detecting influenza in respiratory secretions.
27. Wommack KE, Bhavsar J, Ravel J. Metagenomics: read length matters. *Appl Environ Microbiol*. 2008; 74:1453–1463. [PubMed: 18192407]
28. Yozwiak NL, Skewes-Cox P, Gordon A, Saborio S, Kuan G, Balmaseda A, Ganem D, Harris E, DeRisi JL. Human enterovirus 109: a novel interspecies recombinant enterovirus isolated from a case of acute pediatric respiratory illness in Nicaragua. *J Virol*. 2010; 84:9047–9058. [PubMed: 20592079]
29. Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX, Ding SW. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc Natl Acad Sci U S A*. 2010; 107(4):1606–1611. [PubMed: 20080648] First report of deep sequencing of small interfering RNA for virus identification. Focusing on siRNA enriches for nucleic acid of replicating viruses in insect cells.
30. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res*. 2007; 17:1195–1201. [PubMed: 17600086]
31. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A*. 2001; 98:11609–11614. [PubMed: 11562506]
32. Wommack KE, Williamson KE, Helton RR, Bench SR, Winget DM. Methods for the isolation of viruses from environmental samples. *Methods Mol Biol*. 2009; 501:3–14. [PubMed: 19066805]
33. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. *Nat Protoc*. 2009; 4:470–483. [PubMed: 19300441]
34. Rector A, Tachezy R, Van Ranst M. A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. *J Virol*. 2004; 78:4993–4998. [PubMed: 15113879]
35. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B. Identification of a new human coronavirus. *Nat Med*. 2004; 10:368–373. [PubMed: 15034574]
36. Maan S, Rao S, Maan NS, Anthony SJ, Attoui H, Samuel AR, Mertens PP. Rapid cDNA synthesis and sequencing techniques for the genetic study of bluetongue and other dsRNA viruses. *J Virol Methods*. 2007; 143:132–139. [PubMed: 17433453]

37. Clem AL, Sims J, Telang S, Eaton JW, Chesney J. Virus detection and identification using random multiplex (RT)-PCR with 3'-locked random primers. *Virology*. 2007; 4:65. [PubMed: 17598900]
38. Xiang CC, Chen M, Ma L, Phan QN, Inman JM, Kozhich OA, Brownstein MJ. A new strategy to amplify degraded RNA from small tissue samples for microarray studies. *Nucleic Acids Res*. 2003; 31 e53.
39. Mizutani T, Endoh D, Okamoto M, Shirato K, Shimizu H, Arita M, Fukushi S, Saijo M, Sakai K, Lim CK, et al. Rapid genome sequencing of RNA viruses. *Emerg Infect Dis*. 2007; 13:322–324. [PubMed: 17479903]
40. Watanabe S, Mizutani T, Sakai K, Kato K, Tohya Y, Fukushi S, Saijo M, Yoshikawa Y, Kurane I, Morikawa S, et al. Ligation-mediated amplification for effective rapid determination of viral RNA sequences (RDV). *J Clin Virol*. 2008; 43:56–59. [PubMed: 18595773]
41. Winget DM, Wommack KE. Randomly amplified polymorphic DNA PCR as a tool for assessment of marine viral richness. *Appl Environ Microbiol*. 2008; 74:2612–2618. [PubMed: 18344351]
42. Froussard P. A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nucleic Acids Res*. 1992; 20:2900. [PubMed: 1614887]
43. Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A*. 2002; 99:15687–15692. [PubMed: 12429852]
44. Pirc K, Jebbink MF, Berkhout B, van der Hoek L. Detection of new viruses by VIDISCA. Virus discovery based on cDNA-amplified fragment length polymorphism. *Methods Mol Biol*. 2008; 454:73–89. [PubMed: 19057862]
45. Berthet N, Reinhardt AK, Leclercq I, van Ooyen S, Batéjat C, Dickinson P, Stamboliyska R, Old IG, Kong KA, Dacheux L, et al. Phi29 polymerase based random amplification of viral RNA as an alternative to random RT-PCR. *BMC Mol Biol*. 2008; 9:77. [PubMed: 18771595]
46. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, et al. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A*. 2002; 99:5261–5266. [PubMed: 11959976]
47. Ye J, McGinnis S, Madden TL. BLAST: improvements for better sequence analysis. *Nucleic Acids Res*. 2006; 34:W6–W9. [PubMed: 16845079]
48. Svraka S, Rosario K, Duizer E, van der Avoort H, Breitbart M, Koopmans M. Metagenomic sequencing for virus identification in a public-health setting. *J Gen Virol*. 2010; 91:2846–2856. [PubMed: 20660148]
49. Kapoor A, Victoria J, Simmonds P, Wang C, Shafer RW, Nims R, Nielsen O, Delwart E. A highly divergent picornavirus in a marine mammal. *J Virol*. 2008; 82:311–320. [PubMed: 17942560]
50. Fredricks DN, Relman DA. Infectious agents and the etiology of chronic idiopathic diseases. *Curr Clin Top Infect Dis*. 1998; 18:180–200. [PubMed: 9779355]
51. Fredricks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev*. 1996; 9:18–33. [PubMed: 8665474]
52. Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SW, Hibberd ML, Liu ET, Rohwer F, Ruan Y. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol*. 2006; 4 e3.
53. Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, Shaukat S, Masroor MA, Wilson ML, Ndjanga JB, Peeters M, et al. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J Virol*. 2010; 84:1674–1682. [PubMed: 20007276] Example of highly prevalent viruses of unknown pathogenicity recently discovered in humans and animals samples.
54. Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. *J Virol*. 2010; 84:13004–13018. [PubMed: 20926577]
55. Greninger AL, Holtz L, Kang G, Ganem D, Wang D, DeRisi JL. Serological evidence of human klassevirus infection. *Clin Vaccine Immunol*. 2010; 17:1584–1588. [PubMed: 20739504]
56. Zoll J, Erkens Hulshof S, Lanke K, Verduyn Lunel F, Melchers WJ, Schoondermark-van de Ven E, Roivainen M, Galama JM, van Kuppeveld FJ. Saffold virus, a human Theiler's-like cardiocivirus, is ubiquitous and causes infection early in life. *PLoS Pathog*. 2009; 5 e1000416.

57. Chiu CY, Greninger AL, Chen EC, Haggerty TD, Parsonnet J, Delwart E, Derisi JL, Ganem D. Cultivation and Serological Characterization of a Human Theiler's-Like Cardiovirus Associated with Diarrheal Disease. *J Virol.* 2010; 84:4407–4414. [PubMed: 20164225]
58. Norja P, Hokynar K, Aaltonen LM, Chen R, Ranki A, Partio EK, Kiviluoto O, Davidkin I, Leivo T, Eis-Hubinger AM, et al. Bioportfolio: lifelong persistence of variant and prototypic erythrovirus DNA genomes in human tissue. *Proc Natl Acad Sci U S A.* 2006; 103:7450–7453. [PubMed: 16651522]
59. Manning A, Willey SJ, Bell JE, Simmonds P. Comparison of tissue distribution, persistence, and molecular epidemiology of parvovirus B19 and novel human parvoviruses PARV4 and human bocavirus. *J Infect Dis.* 2007; 195:1345–1352. [PubMed: 17397006]
60. Sharp CP, Lail A, Donfield S, Simmons R, Leen C, Klenerman P, Delwart E, Gomperts ED, Simmonds P. High frequencies of exposure to the novel human parvovirus PARV4 in hemophiliacs and injection drug users, as detected by a serological assay for PARV4 antibodies. *J Infect Dis.* 2009; 200:1119–1125. [PubMed: 19691429]
61. Kean JM, Rao S, Wang M, Garcea RL. Seroepidemiology of human polyomaviruses. *PLoS Pathog.* 2009; 5:e1000363.
62. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science.* 2008; 319:1096–1100. [PubMed: 18202256]
63. Sastre-Garau X, Peter M, Avril MF, Laude H, Couturier J, Rozenberg F, Almeida A, Boitier F, Carlotti A, Coutraud B, et al. Merkel cell carcinoma of the skin: pathological and molecular evidence for a causative role of MCV in oncogenesis. *J Pathol.* 2009; 218:48–56. [PubMed: 19291712]
64. Pancaldi C, Corazzari V, Maniero S, Mazzoni E, Comar M, Martini F, Tognon M. Merkel cell polyomavirus DNA sequences in the buffy coats of healthy blood donors. *Blood.* 2011
65. Shuda M, Feng H, Kwun HJ, Rosen ST, Gjoerup O, Moore PS, Chang Y. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. *Proc Natl Acad Sci U S A.* 2008; 105:16272–16277. [PubMed: 18812503]
66. Blinkova O, Kapoor A, Victoria J, Jones M, Wolfe N, Naeem A, Shaikat S, Sharif S, Alam MM, Angez M, et al. Cardioviruses are genetically diverse and cause common enteric infections in South Asian children. *J Virol.* 2009; 83:4631–4641. [PubMed: 19193786]
67. Kapoor A, Victoria J, Simmonds P, Slikas E, Chieochansin T, Naeem A, Shaikat S, Sharif S, Alam MM, Angez M, et al. A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. *Proc Natl Acad Sci U S A.* 2008; 105:20482–20487. [PubMed: 19033469]
68. Kistler A, Avila PC, Rouskin S, Wang D, Ward T, Yagi S, Schnurr D, Ganem D, DeRisi JL, Boushey HA. Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. *J Infect Dis.* 2007; 196:817–825. [PubMed: 17703411]
69. Lamson D, Renwick N, Kapoor V, Liu Z, Palacios G, Ju J, Dean A, St George K, Briese T, Lipkin WI. MassTag polymerase-chain-reaction detection of respiratory pathogens, including a new rhinovirus genotype, that caused influenza-like illness in New York State during 2004–2005. *J Infect Dis.* 2006; 194:1398–1402. [PubMed: 17054069]
70. Oberste MS, Maher K, Kilpatrick DR, Pallansch MA. Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. *J Virol.* 1999; 73:1941–1948. [PubMed: 9971773]
71. Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochem Biophys Res Commun.* 1997; 241:92–97. [PubMed: 9405239]
72. Simons JN, Leary TP, Dawson GJ, Pilot-Matias TJ, Muerhoff AS, Schlauder GG, Desai SM, Mushahwar IK. Isolation of novel virus-like sequences associated with human hepatitis. *Nat Med.* 1995; 1:564–569. [PubMed: 7585124]
73. Linnen J, Wages JJ, Zhang-Keck ZY, Fry KE, Krawczynski KZ, Alter H, Koonin E, Gallagher M, Alter M, Hadziyannis S, et al. Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science.* 1996; 271:505–508. [PubMed: 8560265]

74. Ninomiya M, Takahashi M, Nishizawa T, Shimosegawa T, Okamoto H. Development of PCR assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. *J Clin Microbiol.* 2008; 46:507–514. [PubMed: 18094127]
75. Hino S, Miyata H. Torque teno virus (TTV): current status. *Rev Med Virol.* 2007; 17:45–57. [PubMed: 17146841]
76. Seifried C, Weber M, Bialleck H, Seifried E, Schrezenmeier H, Roth WK. High prevalence of GBV-C/HGV among relatives of GBV-C/HGV-positive blood donors in blood recipients and in patients with aplastic anemia. *Transfusion.* 2004; 44:268–274. [PubMed: 14962319]
77. Thomas DL, Nakatsuji Y, Shih JW, Alter HJ, Nelson KE, Astemborski JA, Lyles CM, Vlahov D. Persistence and clinical significance of hepatitis G virus infections in injecting drug users. *J Infect Dis.* 1997; 176:586–592. [PubMed: 9291303]
78. Alter HJ, Nakatsuji Y, Melpolder J, Wages J, Wesley R, Shih JW, Kim JP. The incidence of transfusion-associated hepatitis G virus infection and its relation to liver disease. *N Engl J Med.* 1997; 336:747–754. [PubMed: 9052652]
79. zur Hausen H, de Villiers EM. TT viruses: oncogenic or tumor-suppressive properties? *Curr Top Microbiol Immunol.* 2009; 331:109–116. [PubMed: 19230560]
80. Cheval J, Sauvage V, Frangeul L, Dacheux L, Guigon G, Dumey N, Pariente K, Rousseaux C, Dorange F, Berthet N, et al. Evaluation of High Throughput Sequencing for identifying known and unknown viruses in biological samples. *J Clin Microbiol.* 2011 Published online ahead of print on 29 June 2011. Diagnostic sensitivity of deep sequencing measured by spiking biological samples with different type of viral genomes approaches that of PCR.
81. Victoria JG, Wang C, Jones MS, Jaing C, McLoughlin K, Gardner S, Delwart EL. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. *J Virol.* 2010; 84:6033–6040. [PubMed: 20375174] Identification of viral contamination in a biological product using deep sequencing.
82. Neverov A, Chumakov K. Massively parallel sequencing for monitoring genetic consistency and quality control of live viral vaccines. *Proc Natl Acad Sci U S A.* 2010; 107:20063–20068. [PubMed: 21041640]
83. Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A.* 2002; 99:15687–15692. [PubMed: 12429852]
84. Palacios G, Quan PL, Jabado OJ, Conlan S, Hirschberg DL, Liu Y, Zhai J, Renwick N, Hui J, Hegyi H, et al. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg Infect Dis.* 2007; 13:73–81. [PubMed: 17370518]