

Virome genomics: a tool for defining the human virome

Kristine M Wylie^{1,2}, George M Weinstock¹ and Gregory A Storch²

High throughput, deep sequencing assays are powerful tools for gaining insights into virus–host interactions. Sequencing assays can discover novel viruses and describe the genomes of novel and known viruses. Genomic information can predict viral proteins that can be characterized, describe important genes in the host that control infections, and evaluate gene expression of viruses and hosts during infection. Sequencing can also describe variation and evolution of viruses during replication and transmission. This review recounts some of the major advances in the studies of virus–host interactions from the last two years, and discusses the uses of sequencing technologies relating to these studies.

Addresses

¹The Genome Institute, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St. Louis, MO 63108, United States

²The Department of Pediatrics, Washington University School of Medicine, Campus Box 8116, 660 S. Euclid Avenue, St. Louis, MO 63110, United States

Corresponding author: Wylie, Kristine M (kwylie@genome.wustl.edu)

Current Opinion in Microbiology 2013, 16:479–484

This review comes from a themed issue on **Host–microbe interactions: viruses**

Edited by **Carlos F Arias**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 23rd May 2013

1369-5274 © 2013 The Authors. Published by Elsevier Ltd.

Open access under [CC BY license](#).

<http://dx.doi.org/10.1016/j.mib.2013.04.006>

Introduction

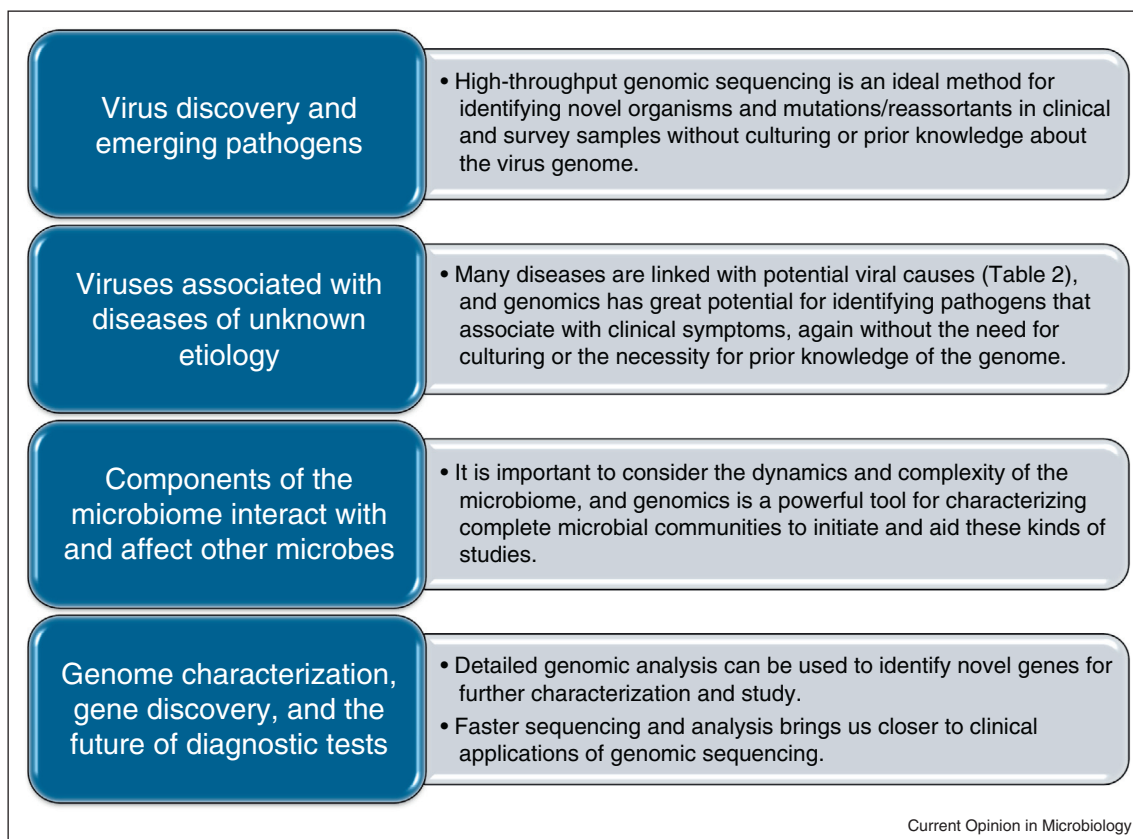
The dynamic interactions between viruses and their hosts during infection are complex. Viruses manipulate the cellular environment and subvert host immune responses in order to replicate, and the host counters the virus's maneuvers in order to control the infection. The interplay between virus and host can be studied in many ways: natural infections; model systems (either animal or cell culture); manipulation of virus–host interactions; identification of the proteins involved in virus–host interactions; and studies of protein functions. High throughput, deep sequencing is a powerful tool for gaining insights into virus–host interactions. Sequencing assays can predict

novel viruses and describe the genomes of novel and known viruses. Genomic information can be used to discover viral proteins that can then be characterized, describe genes in the host that are important in controlling infections, and evaluate gene expression of viruses and hosts during infection. Sequencing can also assess variation and evolution of viruses during replication and transmission. This review recounts some of the major advances in the studies of virus–host interactions from the last two years, and discusses the uses (or potential uses) of sequencing technologies relating to these studies (Figure 1).

Virus discovery and emerging pathogens

In order to understand how viruses interact with their hosts and how they affect human health, we must understand the scope of viral diversity and be able to detect the viruses present in clinical samples. High-throughput, deep sequencing has proven to be an effective tool for this purpose. The relatively unbiased approach it offers for screening clinical samples enables virus discovery without preconceptions about which viruses might be present in the samples. After 10 years of applying this technology to virus detection, eukaryotic virus discovery continues to be robust. A recent example of this is the novel rhabdovirus, Bas-Congo virus, which is an emerging pathogen associated with acute hemorrhagic fever [1], notable for being the first instance of a rhabdovirus being implicated as a cause of hemorrhagic fever. This virus was characterized in the context of a small outbreak, and the presence of antibodies in an asymptomatic caregiver suggested that person-to-person transmission had occurred. Another emerging pathogen, human coronavirus EMC (HCoV-EMC), was recently identified and characterized following an outbreak in the Middle East [2]. This betacoronavirus causes symptoms resembling those of its sister species, SARS coronavirus, including respiratory symptoms and acute renal failure, although HCoV-EMC is most closely related taxonomically to bat coronaviruses. Using modern technologies, the genome of HCoV-EMC was completely sequenced, and assays have been developed to monitor its presence. This virus is particularly interesting because coronaviruses are typically highly restricted to a specific host, but HCoV-EMC can infect cells from primates (human and monkey), swine, and bats (four families) in culture, suggesting that this virus may utilize a receptor shared among these host groups and may be readily transmitted between hosts [3]. These and similar studies demonstrate that continued viral discovery is needed in order to identify and prepare for the effects of emerging viral pathogens on human health. The techniques and technologies are in place for identifying pathogens with similarities to known viruses

Figure 1



The major topics of this review are summarized. Areas of virology research are noted in the blue boxes, and the text in the gray boxes describes how genomics can be used as a tool in that area of research.

(even remote similarities, see [Table 1](#)). The availability of samples and the funding required for the experiments currently bottleneck virus discovery efforts.

Testing samples from affected individuals during outbreaks of diseases of unknown etiology is important for surveillance of pathogens, but it is also important to identify viruses producing symptoms that are mild or subclinical because infection with these viruses may nevertheless have long-term implications for human health. For example, polyomaviruses and papillomaviruses may establish chronic infections. Some of these viruses, including many alpha papillomaviruses and Merkel cell polyoma virus, are associated with cancer [4–5]. In light of their capacity to transform cells, identifying the full range of polyomaviruses and monitoring their presence could ultimately provide insight into the development of some cancers.

Potentially emerging pathogens are of great concern, and influenza pandemics are of particular interest because transmissions between animal reservoirs and humans are observed and the emergence of pandemic strains is expected. Current molecular technologies allow us to

screen for transmission of influenza between animal species and between humans, and to evaluate mutations and quasispecies. In two highly publicized studies, researchers identified mutations that correlated with airborne mammal-to-mammal transmission of the virus, a trait critical for the development of a pandemic [6*,7*]. Researchers used either an H5N1 influenza that originated in birds but had infected humans or a reassortant virus with the avian subtype H5 hemagglutinin and the other seven segments from a 2009 pandemic H1N1 virus. In these studies, viruses were passaged in ferrets, and isolates that had acquired mutations allowing for airborne transmission between ferrets were identified. The strains were sequenced and the mutations that correlated with ferret-to-ferret transmission were identified. The mutations were in the host receptor binding protein hemagglutinin and in the polymerase complex protein basic polymerase 2, and the studies showed that surprisingly few mutations would be necessary for the virus to achieve the capacity for airborne transmission between ferrets. Furthermore, a second study showed that two of the mutations were already circulating among H5N1 strains in birds [8]. These studies provide valuable insight into the plausibility of emergence of an H5N1 pandemic

Table 1**Computational tools used to identify viral sequences, including those with remote sequence similarity to known viruses**

Software	Function	References
IDBA-UD, MetaVelvet	These programs assemble shorter sequences into longer contiguous sequences. Longer sequences can help identification of more remote sequence similarities	[28,29]
BLASTX and TBLASTX	A translated nucleotide sequence is queried against a database of protein sequences or translated nucleotide sequences. This approach is not as sensitive as the searches involving profile Hidden Markov Models (HMMs) and iterative searches (described below), but it is sufficient to identify many viral sequences, including those from many recently discovered novel viruses.	[30]
Realtime Genomics mapx, MulticoreWare mbblastx, Rapsearch2	These three programs are like BLASTX but accelerated to accommodate the large amount of data generated from high-throughput sequencing.	http://www.realtimegenomics.com , http://www.multicorewareinc.com , [31]
HMMER3	A protein query is used to search a profile Hidden Markov Model (HMM) database. Jackhmmer is an iterative alignment program that is part of this package (similar to PSI-BLAST, described below).	[32]
Hhblits	A profile HMM is constructed for the query sequence and then used to query an database of profile HMMs.	[33]
PSI-BLAST	Relatively closely related proteins are identified and conserved amino acid positions are represented in a profile. The profile is used to search a protein database to identify proteins with more remote homologies. The new proteins are added to the profile and the search is repeated iteratively.	[30]
PHI-BLAST	A protein and specific pattern within the protein are used to query a protein database. Proteins that contain the pattern and are similar to the input protein around the pattern are retrieved. CSI-BLAST performs the search iteratively.	[34]
DELTA-BLAST	Searches a database of profile HMMs and identifies the HMM most closely related to the query. The resulting HMM is then used to search a protein database.	[35]

strain for which surveillance could be maintained to anticipate potential outbreaks. It is important to note that the effects of any mutations on the transmission and pathogenesis of influenza in humans are regulated by the genomic context, and the other influenza genes present may play important roles in promoting or limiting the spread of a virus in nature. The controversy generated by this research stemmed from concerns over accidental release from this laboratory or application of the knowledge by bioterrorists. These studies sparked heated debate worldwide, causing some countries to amend their laws and restrictions for research on pathogens with potential as agents of bioterrorism [9]. Some argue that tighter restrictions will slow research in these areas, potentially leaving us less protected from coming pandemics.

Viruses associated with diseases of unknown etiology

There are many diseases of unknown etiology, including Kawasaki disease and chronic fatigue syndrome that have attracted efforts to discover a microbial cause (e.g. see Table 2). XMRV is a retrovirus that was previously associated with chronic fatigue syndrome and prostate cancer. It was discovered in prostate cancer samples, and its complete genome was sequenced [10]. Expanded studies of this virus suggested that it was present in prostate cancer tissue of 40% of patients with a mutation in the RNase L

gene, which encodes an antiviral function, compared to only 1.5% in prostate cancer patients without the mutation [10]. A separate study also showed an association of XMRV with prostate cancers in patients who did not have an RNase L mutation [11]. Subsequently XMRV was also associated with chronic fatigue syndrome [12]. However, these observations were highly controversial, because other studies found no correlation of XMRV with prostate cancer, chronic fatigue syndrome, or a plethora of other diseases of unknown etiologies. The story of this virus has changed dramatically in the last year as additional studies showed that XMRV was a laboratory contaminant that arose from recombination between two proviruses during the passage of prostate cancer tumor cells through mice during the development of cell lines [13*] and ruled out a causal relationship between XMRV and prostate cancer [14]. This story illustrates the caution that must be exercised in the process of virus discovery. Highly sensitive molecular assays may uncover low-level contaminants as well as legitimate pathogens, so carefully constructed controls are critical for accurate interpretation of the data.

Components of the microbiota interact with and affect other microbes

Interactions of viruses with hosts is not only affected by the host genotype but may be influenced by the effects of the bacterial microbiome. This is an area where research

Table 2**Diseases of unknown etiology with a suspected viral cause**

Disease of unknown etiology	Viruses associated	Reference (review article)
Kawasaki disease	Unknown; many have been examined and ruled out	[36]
Multiple sclerosis	Epstein Barr virus (EBV); human endogenous retroviruses (HERV)	[37–40]
Obesity	Human adenoviruses 36, 37 and 5; animal viruses	[41,42]
Rheumatoid arthritis	EBV and HERV	[39]
Systemic lupus erythematosus	EBV and HERV	[39]
Pityriasis rosea	Human herpesvirus 7, possibly human herpesvirus 6; many other viruses have been examined and ruled out	[43,44]
Amyotrophic lateral sclerosis	Unknown	[45]
Brain tumors	Cytomegalovirus (CMV), human 6, herpes simplex viruses 1 and 2	[46]
Inflammatory bowel disease	CMV	[47]
Bell's palsy	Herpes simplex virus 1	[48]
Inclusion body myositis	Retroviruses	[49,50]

is only beginning, but already interesting observations are being made. For example, the retrovirus MMTV and poliovirus have enhanced ability to infect mice in the presence of gut microbes [15,16]. In the case of MMTV, the virus uses bacterially derived LPS to subvert the antiviral response by a mechanism involving the TLR-4-dependent induction of the immunosuppressive cytokine IL-10 [15]. In the case of poliovirus, gut bacteria appear to enhance infectivity by promoting virus attachment to host cells [16]. Likewise, depletion of commensal bacteria in the airways of mice by antibiotic treatment dampened the immune response to influenza virus, likely due to the role of commensal bacteria in developing and regulating the immune system [17•]. These studies suggest that shifts in bacterial communities may be important considerations when trying to understand the implications of viral infections for human health. In each of these cases, mouse models were used, enabling the controlled comparison of viral infection in genetically identical mice, which were germ-free, antibiotic-treated, or colonized with commensal bacteria. Similar studies correlating viral infections in humans with bacterial communities are important, but much more complicated. Nonetheless, metagenomic sequencing analyses have the potential to broadly characterize all of the microbial organisms present in a clinical sample, a capability that can be especially powerful when coupled with patient information. Ultimately, we may be able to identify bacteria and viruses that frequently co-occur or are mutually exclusive and correlate these patterns with clinical symptoms and outcomes. This could yield insights into pathogenesis as well as better diagnostic tools and treatments.

The implications of co-infections are important when evaluating virus–host interactions. For example, the flavivirus GBV-C has been shown to slow the progression of HIV to AIDS [18,19], and herpes simplex virus 1 infection enhances the risk for infection by HIV [20]. Recently, disease progression of pathogenic SIV infection was shown to be associated with a major expansion of the enteric virome [21•]. These initial observations were

made using high-throughput metagenomic sequencing, which identified at least 32 novel enteric viruses in non-human primates. Virus infections may compromise the integrity of the intestinal epithelial lining, allowing the translocation of immunostimulatory molecules from the gut into the blood stream with subsequent systemic immune activation. This study suggests that viruses that co-infect with SIV may have important roles in the pathogenesis of the disease. Metagenomic analyses of samples from people with HIV, sepsis, immunosuppression from transplants, or even more routine illnesses like rhinovirus infections, may reveal co-infections that correlate with disease development or symptoms. Recent PCR-based studies of respiratory samples have found co-infections with multiple viruses in 12–21.7% of children with acute respiratory illness [22–24]. In another study, co-infections were found to be common in patients presenting to the emergency room with fever without a source, based on metagenomic sequencing [25] and PCR assays [26], with as many as five viral genera detected in samples from some subjects [25]. These studies and others indicate viral co-infections are common and worthy of further investigation.

Genome characterization, gene discovery, and the future of diagnostic tests

Genomic sequences of viral strains can be used to predict viral features that are important for viral replication or pathogenesis. For example, based on alignment of influenza A virus segment 3 (PA) from >1000 strains, researchers recently identified a conserved region that leads to a pause and ribosomal frame shift during translation [27•]. This results in a fusion protein called PA-X consisting of the N-terminus of the RNA-dependent RNA polymerase and the newly discovered peptide. This protein appears to have nuclease activity that dampens the host antiviral response. What this study illustrates is that comparative genomics of well-studied viruses can reveal exciting, novel genomic features with implications for virus replication, pathogenicity, and interactions with the hosts. In addition to these kinds of discoveries, the characterizations

of specific viral genotypes or viral variants that associate with virulence can lead to enhanced diagnostic assays that provide information that was not previously available to clinicians. In a new generation of diagnostic tests, rapid sequencing assays could be used not only to detect viruses in patient samples but to provide information about the viral subtype and the presence of virulence genes, which could be used to predict disease severity and outcome. Sequence-based diagnostic tests could also be used to identify the presence of antiviral drug resistance alleles, which would inform the management of some viral infections, including influenza A and cytomegalovirus. Furthermore, ultra deep sequencing (resulting in >100–1000 sequences at each base position of the viral gene or genome) could be used to assess viral variation and quasispecies, which could be important for monitoring emerging pathogens or to development effective vaccines. Sequencing assays aimed at the human transcriptome also have the potential to be developed into clinical tests to measure the host response to viral infection, which may help both with diagnosis and selection of treatment. An additional advantage of diagnostic tests based on high throughput sequencing is that they may be less vulnerable than PCR-based assays to sequence variation in primer-binding or probe-binding regions.

Conclusion

Studies of virus–host interactions are firmly supported by the usage of genome sequencing technologies. A number of applications and advances that have occurred in the last two years are described here, but as sequencing becomes faster and less expensive, the applications to research and diagnostics expand further. Exploratory sequencing analyses can be powerful tools to carry out targeted, hypothesis-driven studies that lead us to better understand pathogenic effects of viruses.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe J-J, Sittler T, Veeraraghavan N, Ruby JG, Wang C *et al.*: **A novel rhabdovirus associated with acute hemorrhagic fever in central Africa.** *PLoS Pathog* 2012, **8**:e1002924.
 2. van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, Osterhaus ADME, Haagmans BL, Gorbalenya AE, Snijder EJ *et al.*: **Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans.** *MBio* 2012:3.
 3. Müller MA, Raj VS, Muth D, Meyer B, Kallies S, Smits SL, Wollny R, Bestebroer TM, Specht S, Suliman T *et al.*: **Human coronavirus EMC does not require the SARS-coronavirus receptor and maintains broad replicative capability in mammalian cell lines.** *MBio* 2012:3.
 4. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Muñoz N: **Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.** *J Pathol* 1999, **189**:12–19.
 5. Feng H, Shuda M, Chang Y, Moore PS: **Clonal integration of a polyomavirus in human Merkel cell carcinoma.** *Science* 2008, **319**:1096–1100.
 6. Herfst S, Schrauwen EJA, Linster M, Chutinimitkul S, de Wit E, Munster VJ, Sorrell EM, Bestebroer TM, Burke DF, Smith DJ *et al.*: **Airborne transmission of influenza A/H5N1 virus between ferrets.** *Science* 2012, **336**:1534–1541.
- This paper and the Imai *et al.* paper below describe studies in which influenza A viruses were manipulated to create viruses that could spread between mammals by air. This furthers our understanding of a viral trait that is necessary for the emergence of pandemic influenza. These studies also generated discussion about dual-use research.
7. Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, Shinya K, Zhong G, Hanson A, Katsura H, Watanabe S *et al.*: **Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets.** *Nature* 2012, **486**:420–428.
 8. Russell CA, Fonville JM, Brown AEX, Burke DF, Smith DL, James SL, Herfst S, van Boheemen S, Linster M, Schrauwen EJ *et al.*: **The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host.** *Science* 2012, **336**:1541–1547.
 9. **News coverage of mutant influenza research and government responses.** *Nature* 2012. (<http://www.nature.com/news/specials/mutantflu/index.html#news>).
 10. Urisman A, Molinaro RJ, Fischer N, Plummer SJ: **Retraction: identification of a novel gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant.** *PLoS Pathog* 2012, **8** <http://dx.doi.org/10.1371/annotation/7e2efc01-2e9b-4e9b-aef0-87ab0e4e4732>.
 11. Schlaberg R, Choe DJ, Brown KR, Thaker HM, Singh IR: **XMRV is present in malignant prostatic epithelium and is associated with prostate cancer, especially high-grade tumors.** *Proc Natl Acad Sci U S A* 2009, **106**:16351–16356.
 12. Lombardi VC, Ruscetti FW, Gupta Das J, Pfof MA, Hagen KS, Peterson DL, Ruscetti SK, Bagni RK, Petrow-Sadowski C, Gold B *et al.*: **Detection of an infectious retrovirus, XMRV, in blood cells of patients with chronic fatigue syndrome.** *Science* 2009, **326**:585–589.
 13. Paprotka T, Delviks-Frankenberry KA, Cingöz O, Martinez A, Kung H-J, Tepper CG, Hu W-S, Fivash MJ, Coffin JM, Pathak VK: **Recombinant origin of the retrovirus XMRV.** *Science* 2011, **333**:97–101.
- This paper proposes the origin of XMRV, which had previously been thought to be an emerging pathogen but was subsequently shown to be an artifact. This study helps us understand the results from previous, controversial studies that had associated XMRV with prostate cancer and chronic fatigue syndrome.
14. Lee D, Gupta Das J, Gaughan C, Steffen I, Tang N, Luk K-C, Qiu X, Urisman A, Fischer N, Molinaro R *et al.*: **In-depth investigation of archival and prospectively collected samples reveals no evidence for XMRV infection in prostate cancer.** *PLoS ONE* 2012, **7**:e44954.
 15. Kane M, Case LK, Kopaskie K, Kozlova A, MacDermid C, Chervonsky AV, Golovkina TV: **Successful transmission of a retrovirus depends on the commensal microbiota.** *Science* 2011, **334**:245–249.
 16. Kuss SK, Best GT, Etheredge CA, Puijssers AJ, Frierson JM, Hooper LV, Dermody TS, Pfeiffer JK: **Intestinal microbiota promote enteric virus replication and systemic pathogenesis.** *Science* 2011, **334**:249–252.
 17. Abt MC, Osborne LC, Monticelli LA, Doering TA, Alenghat T, Sonnenberg GF, Paley MA, Antenus M, Williams KL, Erikson J *et al.*: **Commensal bacteria calibrate the activation threshold of innate antiviral immunity.** *Immunity* 2012, **37**:158–170.
- This paper demonstrates that commensal bacterial affect the immune response to viruses in mice. The role of the microbiota for establishing human health is illustrated, as well as the interactive effects of microbes.
18. Williams CF, Klinzman D, Yamashita TE, Xiang J, Polgreen PM, Rinaldo C, Liu C, Phair J, Margolick JB, Zdunek D *et al.*: **Persistent GB virus C infection and survival in HIV-infected men.** *N Engl J Med* 2004, **350**:981–990.

19. George SL: **Persistent GB virus C infection is associated with decreased HIV-1 disease progression in the Amsterdam Cohort Study.** *J Infect Dis* 2005, **191**:2156-2157 (author reply 2158-2160).
 20. de Jong MAWP, de Witte L, Taylor ME, Geijtenbeek TBH: **Herpes simplex virus type 2 enhances HIV-1 susceptibility by affecting Langerhans cell function.** *J Immunol* 2010, **185**:1633-1641.
 21. Handley SA, Thackray LB, Zhao G, Presti R, Miller AD, Droit L, ●● Abbink P, Maxfield LF, Kambal A, Duan E *et al.*: **Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome.** *Cell* 2012, **151**:253-266.
- This paper demonstrates that the virome in non-human primates infected with pathogenic SIV is largely expanded, and it provides evidence that the expansion of the virome in the gut may affect the integrity of the gut epithelium and associate with progression of disease.
22. Chorazy ML, Lebeck MG, McCarthy TA, Richter SS, Torner JC, Gray GC: **Polymicrobial acute respiratory infections in a hospital-based pediatric population.** *Pediatr Infect Dis J* 2013 <http://dx.doi.org/10.1097/INF.0b013e31828683ce>.
 23. Marcone DN, Ellis A, Videla C, Ekstrom J, Ricarte C, Carballal G, Vidaurreta SM, Echavarría M: **Viral etiology of acute respiratory infections in hospitalized and outpatient children in Buenos Aires, Argentina.** *Pediatr Infect Dis J* 2012, **1827** <http://dx.doi.org/10.1097/INF.0b013e31828683ce>.
 24. Pretorius MA, Madhi SA, Cohen C, Naidoo D, Groome M, Moyes J, Buys A, Walaza S, Dawood H, Chhagan M *et al.*: **Respiratory viral coinfections identified by a 10-plex real-time reverse-transcription polymerase chain reaction assay in patients hospitalized with severe acute respiratory illness – South Africa, 2009–2010.** *J Infect Dis* 2012, **206**(Suppl. 1):S159-S165.
 25. 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.
 26. Colvin JM, Muenzer JT, Jaffe DM, Smason A, Deych E, Shannon WD, Arens MQ, Buller RS, Lee W-M, Weinstock EJS *et al.*: **Detection of viruses in young children with fever without an apparent source.** *Pediatrics* 2012, **130**:e1455-e1462.
 27. Jagger BW, Wise HM, Kash JC, Walters K-A, Wills NM, Xiao Y-L, ● Dunfee RL, Schwartzman LM, Ozinsky A, Bell GL *et al.*: **An overlapping protein-coding region in influenza A virus segment 3 modulates the host response.** *Science* 2012, **337**:199-204.
- In addition to characterizing the effects of an influenza protein that modulates the host response to infection, this paper demonstrates that comparative genomics can reveal novel genes even in well-studied, small viral genomes.
28. Peng Y, Leung HCM, Yiu SM, Chin FYL: **IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth.** *Bioinformatics* 2012, **28**:1420-1428.
 29. Namiki T, Hachiya T, Tanaka H, Sakakibara Y: **MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads.** *Nucleic Acids Res* 2012, **40**:e155.
 30. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
 31. Zhao Y, Tang H, Ye Y: **RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data.** *Bioinformatics* 2012, **28**:125-126.
 32. Eddy SR: **Accelerated profile HMM searches.** *PLoS Comput Biol* 2011, **7**:e1002195.
 33. Remmert M, Biegert A, Hauser A, Söding J: **HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment.** *Nat Methods* 2012, **9**:173-175.
 34. Zhang Z, Schäffer AA, Miller W, Madden TL, Lipman DJ, Koonin EV, Altschul SF: **Protein sequence similarity searches using patterns as seeds.** *Nucleic Acids Res* 1998, **26**:3986-3990.
 35. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL: **Domain enhanced lookup time accelerated BLAST.** *Biol Direct* 2012, **7**:12.
 36. Yim D, Curtis N, Cheung M, Burgner D: **Update on Kawasaki disease: epidemiology, aetiology and pathogenesis.** *J Paediatr Child Health* 2013 <http://dx.doi.org/10.1111/jpc.12172>.
 37. Owens GP, Gilden D, Burgoon MP, Yu X, Bennett JL: **Viruses and multiple sclerosis.** *Neuroscientist* 2011 <http://dx.doi.org/10.1177/1073858410386615>.
 38. Tselis A: **Epstein-Barr virus cause of multiple sclerosis.** *Curr Opin Rheumatol* 2012, **24**:424-428.
 39. Dreyfus DH: **Autoimmune disease: a role for new anti-viral therapies?** *Autoimmun Rev* 2011, **11**:88-97.
 40. Voumvourakis KI, Kitsos DK, Tsiodras S, Petrikos G, Stamboulis E: **Human herpesvirus 6 infection as a trigger of multiple sclerosis.** *Mayo Clin Proc* 2010, **85**:1023-1030.
 41. Atkinson RL: **Viruses as an etiology of obesity.** *Mayo Clin Proc* 2007, **82**:1192-1198.
 42. van Ginneken V, Sitnyakowsky L, Jeffery JE: **Infectobesity: viral infections (especially with human adenovirus-36: Ad-36) may be a cause of obesity.** *Med Hypotheses* 2009, **72**:383-388.
 43. Chuh A, Chan H, Zawar V: **Pityriasis rosea – evidence for and against an infectious aetiology.** *Epidemiol Infect* 2004, **132**:381-390.
 44. Chuh AAT, Chan HHL, Zawar V: **Is human herpesvirus 7 the causative agent of Pityriasis rosea? A critical review.** *Int J Dermatol* 2004, **43**:870-875.
 45. Sica REP, Nicola AFD, Deniselle MCG, Rodriguez G, Monachelli GMG, Peralta LM, Bettini M: **Sporadic amyotrophic lateral sclerosis: new hypothesis regarding its etiology and pathogenesis suggests that astrocytes might be the primary target hosting a still unknown external agent.** *Arq Neuropsiquiatr* 2011, **69**:699-706.
 46. Kofman A, Marcinkiewicz L, Dupart E, Lyshchev A, Martynov B, Ryndin A, Kotelevskaya E, Brown J, Schiff D, Abounader R: **The roles of viruses in brain tumor initiation and oncomodulation.** *J Neurooncol* 2011, **105**:451-466.
 47. Mann EA, Saeed SA: **Gastrointestinal infection as a trigger for inflammatory bowel disease.** *Curr Opin Gastroenterol* 2012, **28**:24-29.
 48. Kennedy PG: **Herpes simplex virus type 1 and Bell's palsy – a current assessment of the controversy.** *J Neurovirol* 2010, **16**:1-5.
 49. Dalakas MC: **Sporadic inclusion body myositis – diagnosis, pathogenesis and therapeutic strategies.** *Nat Clin Pract Neurol* 2006, **2**:437-447.
 50. Machado P, Miller A, Holton J, Hanna M: **Sporadic inclusion body myositis: an unsolved mystery.** *Acta Reumatol Port* 2009, **34**:161-182.