Viral surveillance and discovery
Walter Ian Lipkin and Cadhla Firth

The field of virus discovery has burgeoned with the advent of high throughput sequencing platforms and bioinformatics programs that enable rapid identification and molecular characterization of known and novel agents, investments in global microbial surveillance that include wildlife and domestic animals as well as humans, and recognition that viruses may be implicated in chronic as well as acute diseases. Here we review methods for viral surveillance and discovery, strategies and pitfalls in linking discoveries to disease, and identify opportunities for improvements in sequencing instrumentation and analysis, the use of social media and medical informatics that will further advance clinical medicine and public health.

Addresses
Center for Infection and Immunity, Mailman School of Public Health of Columbia University, 722 West 168th Street, New York, NY 10025, USA

Corresponding author: Lipkin, Walter Ian (wil2001@columbia.edu)

Current Opinion in Virology 2013, 3:199–204
This review comes from a themed issue on Emerging viruses
Edited by Ian Lipkin and Ab Osterhaus
For a complete overview see the Issue and the Editorial
Available online 17th April 2013
1879-6257/$ – see front matter, © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.coviro.2013.03.010

Introduction
The changing dynamics of emerging viruses has resulted in an increasingly important role for viral discovery and diagnostics in clinical medicine and public health. Globalization of travel and trade in domestic animals and animal products, bush meat trafficking, political instability and bioterrorism, as well as climate change and its effects on vector distribution have all facilitated the emergence and reemergence of zoonoses [1–3]. Viruses previously restricted to one host species or geographic region can now appear in unexpected locations, confounding clinicians who are unprepared to recognize new syndromes or to detect new pathogens with their existing diagnostic tests. Immunosuppression associated with organ transplantation, chemotherapy or HIV/AIDS has enabled viruses with limited virulence to become more pathogenic [4–7]. The increased appreciation for the potential role of viruses in chronic cardiovascular, endocrine, neurodevelopmental and neoplastic disorders coupled with a growing repertoire of diagnostic tools and techniques has resulted in a re-energized quest for associated pathogens (Figure 1) [8–10]. This burgeoning interest in discovery and diagnostics now also has practical applications given recent improvements in drugs and vaccines. The lack of effective therapies once made the diagnosis of viral infection primarily an academic exercise; however, the expanding armamentarium of countermeasures for specific viruses promises unprecedented opportunities to reduce morbidity, mortality and the economic costs of viral infections.

Tools for the hunt
Tissue culture
Although this classical method for viral discovery is dismissed by some investigators as cumbersome, expensive and quaint, we continue to use it for several reasons. First, the growth of an agent in culture provides an excellent source of enriched template for molecular characterization. Additionally, having an inoculum for animal model studies, a cell-based assay for serology or a titered stock for neutralization tests greatly aids the pursuit of the proof of disease causation [11–14]. Such stocks are also needed for in vitro assessment of vaccines and to test drug candidates. Finally, visualization of an agent by electron microscopy is an intellectually satisfying coup de grace for even the most jaded microbe hunter [15,16]. The challenge is to find a cell line that allows both entry and replication. The former requires expression of the appropriate receptors (and in some instances, co-receptors), while the latter requires host factors that support virus transcription, replication, assembly and egress [17,18]. Another impediment can be the induction of innate immune responses that impede transcription, replication and/or protein synthesis. Of these challenges, innate immunity is the most straightforward to address, as responses can be depressed through the use of antibodies or other molecules that target alternate pathways. The only current solution for the receptor and factor challenge is to test multiple cell lines, both continuous and primary, from a wide range of host species. Unfortunately, in some instances, these efforts fail and the only system that may prove useful for virus amplification is animal inoculation [17].

Imaging
Until the advent of rapid, inexpensive unbiased high throughput sequencing, immunohistochemistry was frequently the first step towards identifying candidates for molecular assays. In 1999, for example, our decision to pursue flavivirus consensus PCR with brain extract of victims of the encephalitis outbreak in New York was guided by immunoassays performed by Sherif Zaki at the Centers for Disease Control in Atlanta [19–21]. More recently, imaging has been used primarily for examining the strength of the association between an infectious
agent and a disease by testing for its presence at the site of pathology [22–24]. Although rapid protocols for electron microscopy exist that allow the visualization of viruses in a few hours, they require substantial operator expertise and a high agent concentration. Immunohistochemical assays can be pursued using serum from either the affected individual or one with a similar disease, or with antibodies generated in animal models inoculated with the virus candidate or immunized with viral proteins [25]. Where serum is already available these approaches can be implemented rapidly; however, with truly novel agents the delay required to create immunological reagents is typically measured in weeks to months. An alternative is to pursue in situ hybridization. In as little as a few days, genetic sequence of the pathogen candidate can be used to design probes with which to examine the distribution of that candidate in tissues [26].

Multiplex assays
Clinical syndromes are only rarely pathognomonic of infections with specific agents. Thus, although singleplex assays are more common in clinical microbiology, multiplex assays can be more useful not only in differentiating amongst likely candidate pathogens, but also in excluding them, so that efforts can be focused on discovery [27–30]. Multiplex PCR assays in current application detect products using fluorescence, mass spectroscopy or capillary electrophoresis [31–35]. The most highly multiplexed of these can address as many as 50 candidate agents with the sensitivity required for analysis of clinical materials [36–38]. Even higher levels of multiplexing can be obtained with DNA microarrays that can comprise millions of individual probes [39–42]. However, sensitivity may be insufficient for clinical applications. This drawback may ultimately be overcome through the use of nanofluidics and electronic nanocircuitry [43–45*].

High throughput sequencing
The advent of rapid, inexpensive platforms for DNA sequencing has revolutionized microbiology by providing culture-independent methods for pursuing virus discovery and characterization, surveillance and outbreak investigation. Over the past 10 years, the per base cost of sequencing has decreased 10 000-fold, from $5000 per megabase using capillary electrophoresis to $0.5 per megabase using the Illumina platform [46,47]. Accordingly, as the volume of sequence data has grown exponentially in individual experiments and international data repositories, the challenge has shifted from sequence acquisition to analysis. A key determinant in the complexity of bioinformatics
analysis is the read length, which ranges from 150 bp on the Illumina HiSeq or MiSeq instruments to 700 bp on the Roche GS FLX Titanium pyrosequencer [46,48]. The shorter the read length, the greater the number of calculations needed to assemble continuous strings of genomic sequence. An additional challenge on some instruments (e.g. GS FLX and Ion Torrent Personal Genome Machine [PGM]) is the uncertainty of base identification caused by nucleotide homopolymers [46,49-52]. Although single molecule sequencers are in development that promise to provide both fidelity and long strings of sequence data, the current strategy for many investigators is to use the GS FLX or PGM platforms to generate a scaffold for assembling and mapping sequence data from Illumina instruments [53,54**,55**]. Irrespective of platform, the process following sequence acquisition is similar for pathogen discovery: contiguous sequences are assembled, host sequence is subtracted, and the residual sequences are examined for similarity to known microbial sequences at the nucleotide or amino acid level [56,57]. Where no similarities are found, one can examine nucleotide composition and/or order for patterns consistent with viral genera and host species. This approach can also help in deciding which sequences merit further analysis, an important advantage particularly in examining fecal specimens where sequences may represent viruses infecting the host, or a plant or animal consumed by host [58-63]. Although investigations of individual cases or outbreaks of disease still typically focus first on testing for the presence of specific agents using singleplex or syndrome-specific multiplex assays, as the costs of sequencing continues to decrease, instruments become more portable and methods for sample preparation and sequence analysis become more user friendly, we anticipate that sequencing will move to the forefront of clinical diagnosis as well as discovery.

Serology
Although a wide range of genetic methods have been established for multiplex detection of viruses – oligonucleotide microarrays, multiplex PCR and sequencing – serology is largely confined to the singleplex methods that have been employed for decades, including immunofluorescence, enzyme linked immunosorbent (ELISA), Western blot, lateral flow and neutralization assays. Thus, with the notable exception of the Lumexin system wherein individual bead populations are decorated with panels of different antigens that are simultaneously queried, investigators must commit to testing specific pathogen hypotheses using individual tests rather than pursuing syndromic surveillance for a wide range of agents [64]. An additional challenge is that the development of immunoassays is typically a time consuming and labor intensive process that requires preparation of antigen and production or characterization of antisera for validation. It is impossible to estimate rates of morbidity and mortality associated with infection unless one can determine not only the presence of an agent in individuals with disease but also the number of individuals who have been infected but have no or only mild disease. Despite these drawbacks, serology can be useful when molecular methods are unsuccessful. For example, direct means of genetic detection are not useful when viral nucleic acids are not present in an accessible sample (e.g. cerebrospinal fluid in most cases of encephalitis) or where disease may be triggered by an agent that is no longer present (e.g. chronic or degenerative diseases). The risk of exposure to a new infectious agent is modulated by an individual’s previous exposures to similar agents or vaccines. Such exposures may confer complete or partial protection or result in increased risk for more severe disease owing to antibody-mediated enhancement as in dengue [65*]. Thus, knowledge of an individual’s immunological history may influence decisions concerning his/her treatment, vaccination or deployment as a first responder or health care worker in areas with increased probability of encountering specific pathogens. There is urgent need, therefore, to develop serological platforms that can be rapidly deployed as new agents are discovered and to simultaneously detect evidence of exposure to more than one agent. One such platform is Luciferase immunoprecipitation systems (LIPS), a quantitative high-throughput liquid phase format assay wherein viral protein antigens are fused to an enzyme reporter, Renilla luciferase (ReL), expressed in mammalian cells, harvested and then incubated with serum and staphylococcal protein A/G magnetic beads [12,66**]. Following precipitation, luciferase activity is determined as an index to antibody concentration. Establishment of a LIPS assay for a new virus can be developed within several days of the acquisition of genomic sequence data. Another promising platform builds on the microarray technology currently used for detection of the nucleic acids of pathogens [67-69]. Early versions based on spotted peptides have been supplemented by 25 mm × 75 mm arrays onto which peptides are synthesized in situ at densities that exceed 1 million features. This density already allows designs that tile the proteomes of hundreds of viruses. As the printing density increases and efficiencies of scale drive down costs we envision that high throughput multiplex serology will become feasible in clinical as well as research laboratories.

Social media
Recent advances in the portability and accessibility of the Internet affords new opportunities for viral surveillance and discovery. Individuals can rapidly and inexpensively share reports of outbreaks of disease as well as primer sequences, protocols and insights, or request assistance. The first program to build such a presence was ProMED-mail (Programme for Monitoring Emerging Infectious Diseases), created in 1994 to provide a free, interactive international forum for reporting and discussing outbreaks of disease [70]. It remains active to the present
day and serves more than 60,000 subscribers in 185 different countries. Two other popular programs are GPHIN (Global Public Health Intelligence Network) [71] and HealthMap [72]. GPHIN is a subscription news service that reviews media reports and web entries for information concerning outbreaks [71]. HealthMap is a hybrid of the passive and active surveillance strategies used by ProMED-mail and GPHIN, respectively, which provides user-friendly maps of disease emergence with informative hyperlinks [72]. Although none of these programs yet integrates the full range of data sources that have the potential to contribute to global awareness (e.g., medical or veterinary service use, purchases of prescription drugs and over-the-counter remedies), they have dramatically enhanced communication amongst scientists, clinicians, public health practitioners, and laid the foundation for a global warning system that enables focused investments in surveillance and discovery.

Mechanisms of pathogenesis and proof of causation

In the most straightforward examples of pathogen discovery the investigator encounters an outbreak of what appears to be a transmissible disease. Many individuals present with similar signs and symptoms and there is: (1) a pattern consistent with either direct human-to-human transmission or exposure to a common source whether aerosol, enteric or arboviral; (2) a high concentration of the agent at the site of pathology; (3) an acute presentation of disease. However, limiting microbe hunting to such ideal scenarios would eliminate opportunities for actionable insights into chronic or sporadic disorders or early zoonotic transmission events. Host factors including age, nutrition, genetic context, co-infection, and previous exposures to related agents can have a profound impact on susceptibility to infection and the expression of disease. Viruses can cause damage at the site of infection as a direct result of replication, or as an indirect effect of innate or adaptive immune responses to microbial gene products. They may also induce neoplasia through cell cycle interference. The gold standard for proof of causation are the criteria established by Loeffler and Koch known as Koch’s postulates [73]. They require that an agent be present in every case of the disease, specific for the disease and sufficient to reproduce the disease after culture and inoculation into a naïve host. Others including Rivers, Fredericks and Relman, and the current authors have proposed alternative criteria to allow for instances where an agent cannot be propagated in culture or there is no suitable animal model with which to test for transmissibility of disease as well as infection [74–76]. Detailed strategies for pursuing causation are described in recent reviews [56,77]. They rely upon demonstrating strength of association through assessment of adaptive immunity to the candidate agent, quantitation of viral burden, demonstration of footprints of the agent at the site of pathology, biological plausibility by analogy to disease associated with similar agents in the same or other hosts, and prevention or mitigation of disease through use of candidate-specific drugs or vaccines (Figure 1).

Conclusions

The introduction of new tools for surveillance, diagnosis and discovery is timely given the increasing globalization of travel and trade and other factors that enhance the emergence and re-emergence of pathogenic viruses. These tools will have application beyond control of acute disease. We anticipate that they will lead to insights into the pathogenesis of a wide range of idiopathic conditions not typically ascribed to infection such as neurodevelopmental, autoimmune and neoplastic disorders. Concomitant advances in methods for vaccine and drug development suggest the potential for dramatic improvements in health and welfare enabled by enhanced viral intelligence.

Acknowledgements

Work in the Center for Infection and Immunity is supported by the National Institutes of Health (AI05718 and AI079231), USAID PREDICT, and the Defense Threat Reduction Agency. The authors thank Meera Bhat and Nsikan Akpan for their assistance.

References and recommended reading

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

- of special interest
- of outstanding interest


Emerging viruses

   • Akeson M: Automated forward and reverse ratcheting of DNA in a
     This paper is one of the first to demonstrate that sequence
     information can be obtained by passing a DNA molecule through a
     protein nanopore, using Ph209 DNA polymerase to slow the speed of
     translocation. This work greatly advanced the use of nanopore
     technology for DNA sequencing..

46. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Garibian SE,
   Wain J, Pallen MJ: Performance comparison of benchtop high-
   throughputs sequencing platforms. Nat Biotechnol 2012,
   30:434-439.

47. Kircher M, Kelso J: High-throughput DNA sequencing –


49. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C,
    Jaffe DB, Nusbaum C, Gehrke A: Analyzing and minimizing PCR
    amplification bias in Illumina sequencing libraries. Genome

50. Kieleczawa J, Mazaika E: Optimization of protocol for

51. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ:
    Amplification-free Illumina sequencing-library preparation
    facilitates improved mapping and assembly of (G+C)-biased

52. Dohm JC, Lottaz C, Borodina T, Himmelbauer H: Substantial
    biases in ultra-short read data sets from high-throughput DNA

53. Chan JZ, Pallen MJ, Oppenheim B, Constantinidou C: Genome
    sequencing in clinical microbiology. Nat Biotechnol 2012,
    30:1068-1071.

54. Croucher NJ, Harris FR, Fraser C, Quail MA, Burton J, van der
    • Linden M, McGee L, von Gottberg A, Song JH, Ko KS et al.: Rapid
      pneumococcal evolution in response to clinical interventions.
      Science 2011, 331:430-434.

This paper demonstrates the power of full genome sequencing to illu-
minate the response of pathogens to selection from host immune pres-
sures or clinical interventions. By sequencing more than 200 complete
organisms of a single strain of multi-drug resistant Streptococcus pneu-
moniae, the authors demonstrate the importance of widespread recom-
bination for rapid adaptation.

55. Rohde H, Qin J, Gui Y, Li D, Loman NJ, Hentschke M, Chen W,
    • Pu F, Peng Y, Li J et al.: Open-source genomic analysis of
      365:718-724.

This paper describes how unbiased high throughputs sequencing and
crowd-sourced analysis were used to determine the sequence informa-
tion of the E. coli strain responsible for the 2011 German outbreak. This
work illustrates how the increasing speed and decreasing cost of UHTS
has transformed outbreak response efforts.

56. Mokili JL, Rohwer F, Dutilh BE: Metagenomics and future

57. Tang P, Chiu C: Metagenomics for the discovery of novel

58. Bruzskiewicz E, Thurer A, Schuldes J, Leimbach A,
    Liesegang H, Meyer FD, Boelter J, Petersen H, Gottschalk G,
    Daniel R: Genome sequence analyses of two isolates from the
    recent Escherichia coli outbreak in Germany reveal the
    emergence of a new pathotype: Enteroto-Aggregative-
    Haemorrhagic Escherichia coli (EAHEC). Arch Microbiol 2011,
    193:883-891.

59. Carroll IM, Ringel-Kulka T, Siddle JP, Klaenhammer TR, Ringel Y:
    Characterization of the fecal microbiota using high-
    throughputs sequencing reveals a stable microbial community

60. Hagen C, Frizzi A, Gabriels S, Huang M, Salati R, Gabor B,
    Huang S: Accurate and sensitive detection of gemivirus through
    enrichment, high-throughput sequencing and automated sequence

61. Lamendella R, Domingo JW, Ghosh S, Martinson J, Oerther DB:
    Comparative fecal metagenomics unveils unique functional
    capacity of the swine gut. BMC Microbial 2011,
    11:103.

    plasticity of the bovine abomasal microbiota in immune
    animals in response to Ostertagia ostertagia infection. PLoS
    One 2011, 6:e24417.

63. Shan T, Li L, Simmonds P, Wang C, Moesaer A, Delwart E: The
    fecal virome of pigs on a high-density farm. J Virol 2011,
    85:11697-11708.

64. Kunic N, Merante F, Yaghoubian S, Hismworth D, Janeckzo R:
    Advances in the diagnosis of respiratory tract infections: role
    of the Lumines XTAG respiratory viral panel. Ann N Y Acad Sci

65. Rotham AL: Immunity to dengue virus: a tale of original
    • antigenic sin and tropical cytomeglovirus. Nat Rev Immunol

This paper provides an excellent review of the complex nature of the
immune response to dengue infection. It includes an in-depth discussion of
the effects of cytomeglovirus, T cell response and antibody-dependent
enhancement on the severity of infection and vaccination strategies.

66. Burbelo PD, Bren KE, Ching KH, Gogineni ES, Kottillit S, Cohen JJ,
    • Kovacs JA, Iadarola MJ: LIPS arrays for simultaneous detection
      of antibodies against partial and whole proteomes of HCV, HIV

With the creation of Luciferase immunoprecipitation systems (LIPS),
serological identification of infectious agents became a rapid (~10 days
from the generation of a small fragment of genomic sequence to a
working assay) and highly sensitive technique that could be applied to
any host organism. This paper demonstrates that LIPS assays can simulta-
nuously profile antibody response to a broad panel of antigens.

67. McLoughlin KS: Microarrays for pathogen detection and

68. Panicker RC, Huang X, Yao SQ: Recent advances in peptide-
    based microarray technologies. Comb Chem High Throughput
    Screen 2004, 7:547-556.

69. Robinson WH, DiGennero C, Hueber W, Haab BB, Kamachi M,
    Dean EJ, Fournel S, Fong D, Genoves MC, de Veygar HE et al.: Autoantigen
    microarrays for multiplex characterization of autoantibody responses.

70. Madoff LC: ProMED-mail: an early warning system for

71. Mykhaylovsksiy E, Weir L: The Global Public Health Intelligence
    Network and early warning outbreak detection: a Canadian
    contribution to global public health. Can J Public Health
    2006, 97:42-44.

72. Freifeld CC, Mandl KD, Reis BY, Brownstein JS: HealthMap:
global infectious disease monitoring through automated
classification and visualization of Internet media reports. J Am

73. Koch R: Ueber bakteriologische Forschung, Verhandl. des X.
    Interna, Med. Edited by Berlin, Germany: Cong, Berlin 1890; 1891
    [August Hirschwald (Series Editor)]

74. Fredericks DN, Relman DA: Sequence-based identification of
    microbial pathogens: a reconsideration of Koch’s postulates.

75. Lipkin WI: Microbe hunting. Microbiol Mol Biol Rev 2010,
    74:363-377.

    33:1-12.

77. Lipkin WI: The changing face of pathogen discovery and