Collectively, we face a future where emerging viral pathogens will become more prevalent, fueled by advances in medical therapies, changes in the environment, globalization and better disease surveillance. Modern immunosuppressive therapies, targeting neoplastic and certain chronic diseases, have created new opportunities for viruses that are normally nonpathogenic. As the barriers between animals and humans are removed by the expansion of the human population, and global warming extends the range of certain vectors and pathogens, new viruses will be regularly introduced into the human population. Compounding these forces is globalization, which facilitates the rapid spread of viruses from one continent to another. Recent examples include the West Nile Virus, which has emerged in North America [1]; the SARS coronavirus, which triggered a worldwide epidemic [2–4]; and the swine-origin influenza A/H1N1 virus, which has become the latest pandemic virus [5–9].

In addition, there exist many common clinical syndromes, such as respiratory, gastrointestinal and encephalitic infections, in which many of the causative agents are believed to be viruses but extensive conventional diagnostic tests have been unable to identify specific pathogens [10–15]. Finally, the specter of bioterrorism is ubiquitous, as the knowledge and tools for creating novel, more effective biological agents become more widespread [16–19].

Enhancements in our disease surveillance networks and techniques have allowed us to rapidly identify more instances of emerging infectious diseases [20–24]. However, the traditional paradigm of finding the unknown causes of these diseases relies upon diagnostic tests for known agents. This makes it difficult and inefficient, and sometimes impossible, to identify unexpected or novel viral pathogens using conventional methods. In contrast, more comprehensive metagenomics techniques, such as microarrays and high-throughput sequencing, are ideally suited for the task of systematic virus discovery.

Although conventional methods for virus detection, such as virus culture, electron microscopy, serology and PCR, have been used successfully for identifying new viruses, these methods each have limitations for systematic virus discovery. Many viruses cannot be amplified in cell culture, or will not exhibit characteristic cytopathic effects during their growth [25]. Electron microscopy is a relatively insensitive method for virus detection and provides only morphologic clues regarding the identity of the virus [26]. Sera from previously infected hosts can be used to label viruses to enhance detection in cell culture or electron microscopy, but sera containing high titers of specific viral antibodies can be difficult to obtain [26,27]. PCR targeting conserved genetic regions can be used to detect variants of known viruses, but may not be able to detect more divergent or completely novel viruses for which there exists no a priori sequence data [28]. These conventional tests can be combined and run together, or in series, to detect a wider range of pathogens, but this approach can be costly, inefficient and time consuming [29,30].
Metagenomics strategies to virus discovery typically employ an algorithmic approach (Figure 1). Singleplex PCR assays run in parallel, or multiplex PCR assays, can be used to immediately screen for the most common viral pathogens associated with the infectious disease being studied [31,32]. The resulting PCR products from the various pathogens are resolved using gel electrophoresis, differentially labeled probes or microarrays to identify the virus present in the sample [33–35]. Samples that test negative for the common viruses are then tested by a broad-range virus detection assay, such as a pan-viral microarray, which is capable of detecting viruses that are formally represented on the microarray, as well as novel variants of these viruses [36–38]. Any viruses detected by these methods can then be fully sequenced to detect potential variants of known viral species [39]. If viruses are not detected by PCR or microarray, samples are then subjected to high-throughput sequencing to detect novel viruses that do not have significant sequence homology to any known viruses [40,41]. Prevalence and association studies can be performed to link novel viruses identified by microarray or high-throughput sequencing with disease [42]. Based on the current costs for PCR, microarray analysis and high-throughput sequencing.

**Figure 1. Algorithmic approach to novel virus discovery.** Clinical specimens that are suspected to contain viruses, but which test negative for pathogens by conventional microbiology tests, are screened for common viruses through PCR. If common viruses are present, these are sequenced to determine whether they are novel variants and to facilitate molecular epidemiology investigations. Specimens which are negative for common viruses are hybridized to a pan-viral microarray to detect a broader range of known and unknown viruses. If viruses cannot be found through microarray techniques, then the sample is subjected to high-throughput sequencing to search for viral sequences. Any viral signatures detected by microarray or high-throughput sequencing will lead to attempts to recover the entire virus genome sequence in order to better characterize the novel virus.
sequencing, this algorithmic approach to virus detection and discovery allows for the systematic detection of a potentially unlimited range of viruses in the most cost-effective, yet comprehensive, manner. As the relative costs of these tests change and new technologies are made available, the algorithm will naturally evolve to reflect these changes.

**New viral diseases**

The first clue to a new viral disease depends upon an astute clinician recognizing an unusual pattern of signs and symptoms that do not match those of known infectious diseases [43]. Alternatively, there may be a patient with a previously recognized clinical syndrome for which laboratory tests are negative for the known etiologic agents. Clusters of a new infectious disease may be recognized by epidemiologists using traditional or novel surveillance tools, such as analyzing worldwide news feeds, internet search-engine trends and open-source real-time surveillance systems [20–24]. Active laboratory surveillance of human, animal and environmental samples may detect new or unusual viruses, which will lead to additional studies to determine their role in human health [44–47]. From specimens submitted for diagnostic testing, the clinical microbiology laboratory may identify unusual laboratory findings that indicate the presence of a novel virus. There are also large collections of banked samples in most clinical laboratories for which conventional tests were unable to identify a pathogen, but which potentially harbor novel viruses.

In the search for new viral diseases, the safer hunting grounds include many acute infectious diseases of unknown etiology [43]. The probability of finding new viruses in diseases, such as acute gastroenteritis, respiratory tract infections, meningitis and encephalitis, is high [10–15]. Many viruses are known to cause these syndromes, and the remaining fraction of cases of unknown etiology may also be largely attributed to as yet undiscovered viruses. As our techniques for virus discovery improve, we will enhance our ability to target viruses in the more perilous hunting grounds of idiopathic, neoplastic, autoimmune, inflammatory, endocrine and other chronic diseases, where the associations between pathogen and disease are more difficult to establish [43,48–52]. Environmental and host genetic factors may also play a significant role in the pathogenesis of these diseases, and complex studies may be required to ascertain whether virus infection is the primary cause or whether the etiology of the disease is multifactorial [53].

Other efforts to identify new viral diseases target animals as sources of new and emerging human viruses [45,46]. Monitoring for novel viruses at the animal–human interface could provide an early warning system for new infectious diseases before they spread further into the human population. It is likely that a large pool of novel human viruses still exists owing to both viral evolution and undiscovered viruses in animals, thus supporting the role for active virus discovery programs [46,54].

The scarcity of effective antiviral therapies has been one major argument against the utility of efforts to discover new viral agents. However, it is clear that to begin development of new antiviral therapies, the viral etiologies of many idiopathic diseases must first be uncovered and the epidemiology of these new viruses understood [55]. Furthermore, even in the absence of effective antiviral drugs, it may be possible to discontinue unnecessary antibiotics and avoid additional diagnostic tests once the etiologic agent is identified, thus reducing adverse effects from unneeded drugs and reducing overall healthcare costs [56,57]. The identification of a novel virus also allows for the development of new viral detection and serologic assays to better understand the epidemiology and pathogenesis of the new infectious disease [58–60]. Finally, the expectations of clinicians and patients are greater as the tolerance for diagnostic uncertainty decreases in an age where information is both widely available and immediately accessible [61].

**Traditional strategies**

Although new metagenomics techniques promise to dramatically increase our rate of virus discovery, traditional laboratory techniques in virology have served well for virus discovery, even in modern times. The cornerstone of virus discovery has been the use of cell culture techniques, since the amplification of the virus in cell culture facilitates downstream analysis with other laboratory methods, including electron microscopy, serology and PCR [62]. These other methods often complement the use of virus culture, providing additional evidence that the novel virus is the etiologic agent rather than a tissue culture contaminant or a transient virus unrelated to the disease [27,65,64]. The electron microscope has made important contributions to virus discovery, from historically being the primary method of virus detection in clinical samples, such as in the discovery of Norwalk virus [65,66], to a rapid tool for identifying new viruses in cell culture, such as in the initial identification of
SARS coronavirus, Melaka and Saffold viruses from infected cell cultures displaying cytopathic properties [4,67,68]. With morphologic information from electron microscopy, it is then possible to narrow the identity of the virus into select taxonomic groups, thus allowing for the use of PCR targeting conserved genetic regions within these groups of viruses [69,70]. Serologic methods have been combined with electron microscopy, as was the case in the discovery of hepatitis B virus, where serum from infected patients was used to label virus particles for immuno-electron microscopy [71,72]. Immune serum has also been used to screen peptide expression libraries generated from samples infected with an unknown virus; the classic application of this method was the discovery of hepatitis C virus from patients with non-A, non-B hepatitis [73]. Nevertheless, the systematic use of these traditional methods for virus discovery is problematic, as they are often labor intensive and lack sufficient analytical sensitivity compared with modern molecular detection techniques.

Viral nucleic acid extraction

Perhaps the most crucial step in the use of molecular techniques for virus discovery is the efficient extraction of viral nucleic acid from the clinical specimen [74–76]. Both the absolute and relative concentrations of viral relative to nonviral nucleic acids determine the success of subsequent molecular assays. For clinical specimens that are solid or viscous in nature, homogenization or enzymatic digestion of the specimen is necessary to liberate virus particles for extraction [77–79]. Various filtration and differential centrifugation techniques can be used to further concentrate virus particles [29,80,81]. The genomes of intact virus particles are protected by nucleocapsids, and endonuclease treatment can be used to remove unprotected RNA and DNA also present in the sample [82]. All of these steps increase the absolute and relative concentration of viral nucleic acids prior to nucleic acid extraction. Many manual and automated methods are available for nucleic acid extraction and the selection of the best method will need to balance extraction efficiency and throughput [83,84]. Unfortunately, viruses which exist in episomal forms or that have integrated into the host genome will not be readily amenable to these purification strategies. Additional methods to reduce the molecular complexity of such nucleic acid samples include selection for polyadenylated RNA [85]. The concentration of viral nucleic acids and reduction of sample complexity serve to increase the sensitivity of nucleic acid amplification tests, subtraction methods, microarrays and high-throughput sequencing for novel virus discovery.

Nucleic acid amplification tests

Nucleic acid amplification tests, such as PCR, have traditionally been used for the detection of known viruses. The use of PCR requires a priori knowledge of the virus genome and, thus, precludes its use for the discovery of viruses whose genomes are highly divergent from that of known viruses [86]. However, novel variants of known viral species can be discovered through PCR detection followed by genome sequencing. New types of coronaviruses, herpesviruses, enteroviruses, parechoviruses and rhinoviruses have been identified in this way [87–92]. Designing degenerate or specific PCR primers targeting conserved targets within each viral taxon enhances the ability to detect more members of that group of viruses [93]. For example, the consensus-degenerate hybrid oligonucleotide primers (CODEHOP) methodology employs degenerate primers that target stable amino acid sequence motifs, while taking in account redundancies in codon sequences [28,94].

In the algorithmic approach to virus discovery (Figure 1), PCR is useful for detecting the most common viral causes of a particular clinical syndrome. Samples already containing a common viral pathogen known to cause that syndrome are less likely to also yield additional novel viruses and, in most cases, can be excluded from further analysis by pan-viral microarray or high-throughput sequencing [39,41]. In particular, multiplex PCR assays are a cost-effective and efficient way to detect common viral pathogens [31]. Several methods can be used to resolve the amplicons produced by multiplex PCR, including electrophoretic separation, multiple fluorescent probes, bead-based probes, mass spectroscopy and microarray [33–35,95–97].

Differential display & subtraction techniques

Methods to study differential gene expression are also applicable for the discovery of new viral pathogens [86,98–100]. In general, these methods require infected and uninfected samples, such that common genetic sequences between the two sets of samples are ignored or removed. For optimal efficiency, the two sets of samples must be highly related, such as being the same tissue type or derived from the same individual, and essentially identical, except for the presence or absence of the unknown virus [101]. In samples
Two of the more widely used pan-viral microarrays for viral discovery and diagnostics are the Virochip and the GreeneChip [36–38]. Both designs employ 70-mer oligonucleotide probes that target conserved genetic regions within each taxonomic group of viruses, for which sequence data is available. The increased length of the 70-mer oligonucleotide probes makes them more tolerant of mismatches and, thus, more suitable for the detection of novel viruses [36]. A large number of probes are required to provide redundancy and to generate nuances in hybridization patterns that differentiate between known and unknown viruses. The GreeneChip takes advantage of the Protein Families database of alignments in GenBank to generate viral probes, whereas the Virochip utilizes evolutionarily conserved sequences in the viral genomes represented in GenBank [36,107]. In addition, the third and fourth (current) iterations of the Virochip also employ oligonucleotide probes that are specific to individual virus species and subtypes, thus allowing for the enhanced identification of viruses at the species and subspecies levels, as well as increasing its power to differentiate novel from known viruses [108,109]. As new viruses are sequenced and added to the database, the probes on the microarray are regularly updated to reflect the new taxonomic landscape.

After the extraction of nucleic acids, reverse transcription and random amplification are used to amplify the genetic material in the clinical sample. For pan-viral microarrays, random PCR amplification is commonly used, although linear amplification strategies (e.g., PhiX29 strand-displacement or T7-based amplification) may enable quantitative viral detection [110,111]. The use of random amplification creates a problem for the detection of rare viruses when the amount of nonviral nucleic acids is high, such as the detection of viruses in tissue samples. To address this limitation, numerous preanalytical strategies have been employed to increase the virus concentration prior to random amplification. These include physical and enzymatic methods to separate viruses from human cells and bacteria, as outlined in the previous section. In diseases and samples where the concentration of viruses is high relative to other cells, such as respiratory secretions in acute respiratory infections or stool in acute gastroenteritis, the clinical sensitivity of the viral microarrays is comparable to that of specific PCR assays for the same viruses [112]. The use of other target amplification strategies for broad-spectrum microbial detection is covered in greater detail in an article by Leski et al. in this issue [113].
The large number of probes, along with the occurrence of both predicted and unpredicted cross-hybridization, necessitates the use of computer algorithms and statistical methods to identify virus hybridization patterns on the microarray. One computational strategy is E-Predict, an algorithm that compares observed hybridization patterns on the Virochip to a database of theoretical virus hybridization patterns \[114\]. Theoretical hybridization patterns are generated through \textit{in silico} determination of the free energy of hybridization of known viruses to each probe on the microarray. The statistical significance of each virus prediction can be calculated from the theoretical hybridization patterns, and modified in a Bayesian fashion from past cumulative microarray data. To enhance the sensitivity of detection, the statistical significance of the observed intensity of each oligonucleotide probe can also be calculated in reference to a set of control and historical samples. The ranking and taxonomic relationships of the most statistically significant oligonucleotides can then be used to determine the identity of the virus \[57\]. Similar algorithms also exist for determining the significance of pathogen detection microarray data, such as PhyloDetect \[115\] and DetectiV \[116\].

The Virochip has been used in the identification of the SARS coronavirus and for recovery of the first kilobase of genomic sequence \[37,117\]. Subsequently, it has detected many novel viruses including:

- A murine retrovirus associated with prostate cancer \[58\]
- A new clade of rhinoviruses \[112\]
- An avian bornavirus causing proventricular dilatation disease in birds \[109\]
- A coronavirus in a beluga whale \[118\]
- A novel human cardiowirus \[108\]

Both the Virochip and Greenechip have been demonstrated to identify a wide range of viral respiratory pathogens \[10,112\], including cases of bronchiolitis and pneumonia in critically ill patients after extensive conventional laboratory testing failed to yield a diagnosis \[56,57\].

**Sequence-independent amplification**

Sequence-independent amplification methods enable the unbiased detection of highly divergent or novel viruses that cannot be identified by other techniques. The general algorithm begins with nucleic acid amplification using various methods, including randomly primed PCR or restriction endonuclease digestion of the sample DNA and ligation of adaptor sequences, followed by PCR amplification. RNA can be analyzed with a reverse transcription step prior to amplification. After sequence-independent amplification, subcloning and sequencing of the amplified fragments, homology searches in GenBank are conducted to identify sequences that possess similarities to known viruses. Fragments that do not match any known sequence, or that are distantly related to known viral sequences, are of particular interest as they may correspond to novel viruses. Traditionally, sequence-independent amplification methods are best applied to clinical samples, such as body fluids (e.g., respiratory secretions, stool, blood, cerebrospinal fluid and urine). These samples can be highly purified for viruses by filtration, ultracentrifugation, or endonuclease treatment, whereby free nucleic acids are degraded but viral nucleic acids remain protected within the proteinaceous capsid. Examples of sequence-independent amplification methods that have been used for virus discovery include sequence-independent single primer amplification and the adaptation of these techniques for high-throughput sequencing. Sequence-independent amplification methods have been successful in detecting many novel viruses over the past decade, including parvoviruses in human sera \[119\], human metapneumoviruses, bocaviruses and polyomaviruses in respiratory secretions \[120–123\], and human adeno-, parvoviruses, parechoviruses, bocaviruses, cosaviruses and kobuviruses in stool \[124–128\].

**High-throughput sequencing**

The development of high-throughput second-generation sequencing technology has greatly impacted the field of virus discovery by enabling ‘deep’ sequencing of clinical samples with the production of hundreds of thousands to millions of sequence reads per run \[40,129–131\]. This is a sophisticated yet ‘brute force’ method to detect sequences corresponding to novel viruses that may be present at exceedingly low titers in clinical specimens, or may be too divergent from known viruses to be detected by PCR or microarray techniques \[40,132\]. High-throughput sequencing is a rapidly evolving field, and several new third-generation sequencing systems are already in development \[129,133\]. The two platforms most commonly used for virus discovery applications include the pyrosequencing system of 454 Life Sciences/Roche (CT, USA) and the ‘sequencing-by-synthesis’ system of Solexa/Illumina (CA, USA) \[134,135\].

For the detection of novel viruses, critical parameters of any high-throughput sequencing platform include the average read length
and total number of sequence reads generated. The sequences must be long enough to permit unique identification of sequences corresponding to novel viral genomes, and there must be enough reads generated for adequate sequencing depth. Although high-throughput sequencing methods provide unparalleled depth of sequencing, strategies to reduce host background sequences and purify viral nucleic acid sequences remain essential for detection sensitivity. One example is the recent discovery of a novel polyomavirus associated with Merkel cell carcinoma by massively parallel pyrosequencing [132]. Despite selection for polyadenylated RNA, and analysis of over 400,000 sequences generated from tissue biopsy specimens, only two sequences were found to correspond to the novel polyomavirus. The Roche Genome Sequencer FLX™ can currently generate up to 1 million sequence reads per run with an average read length ranging from 200 to 450 bp, while the Illumina Genome Analyzer II can generate up to 80 million sequence reads per run, with fixed paired-end read lengths of 100–140 bp (paired-end reads are sequences generated from the opposite ends of the same individual DNA molecule). Samples can also be barcoded so that multiple samples can be analyzed per run with proportionally fewer sequences generated per specimen [136,137].

The analysis of high-throughput sequencing data can be computationally challenging. A customized bioinformatics pipeline used in our laboratory for virus discovery from high-throughput sequencing reads is shown in Figure 2 [131,135]. Raw sequence reads are trimmed to remove low-complexity and duplicate reads and classified by sample barcode. Overlapping reads are computationally assembled and the resulting unique contiguous sequences (contigs) are subjected to similarity searches to nonredundant organism sequence databases. Sequence reads corresponding to human, bacterial and other nonviral organisms are computationally subtracted from the dataset. The remaining sequence reads are then compared with viral databases, using programs that examine homology, at first at the nucleotide and then at the amino acid levels (BLASTn, BLASTx and tBLASTx). Detected viral sequences corresponding to known viruses are taxonomically classified, and attempts are made to assemble viral contigs or genomes from sequences corresponding to these novel viruses. Finally, sequences that do not align to any viral or nonviral database are examined using a remote homology detection program (such as PHI-BLAST) for the presence of extremely divergent viruses [138].

Recently, the use of high-throughput sequencing technology has led to the identification of several novel viruses – an arenavirus related to lymphocytic choriomeningitis virus in transplant recipients [40], a hemorrhagic fever virus from South Africa [134], a human kobavirus [135] and the aforementioned Merkel cell polyomavirus [132]. Given the high costs and lengthy times for sample preparation and data analysis, unbiased high-throughput sequencing, at this point, is primarily used for clinical samples that test negative for known viral agents by all other methods. In the future, with rapid advances in technology, decreased sequencing costs and sample multiplexing, high-throughput sequencing may be a powerful tool, not only for viral discovery, but also for viral diagnostics in the clinical as well as research settings.

**Figure 2. High-throughput sequencing pipeline for virus discovery.** From the raw sequence data, low complexity and duplicate reads are first removed and any sample barcoding is resolved. *De novo* assembly of overlapping reads is performed, followed by subtraction of all nonviral sequences. The remaining sequences are aligned to existing viral sequence databases to classify and assemble novel viral sequences. Other methods are used to interpret the ‘no hits’ sequences, which have no similarity to either viral or nonviral sequences.
Evidence of causation
The discovery of a novel virus in a clinical sample from an individual with an acute or chronic disease does not imply causation or even *bona fide* association of the virus with the disease. Much of the difficulty in assigning an etiologic role to newly discovered viruses comes from the fact that viruses routinely fail Koch’s postulates for causality [49]. Many viruses cannot grow in culture or else have a restricted host range, thereby making it impossible or unethical to attempt experimental human infection and re-isolation of the virus. Alternatives to Koch’s postulates include serological analysis, in which the identification of specific antibodies to the novel virus may imply a causal relationship, or

<table>
<thead>
<tr>
<th>Executive summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>New viral diseases</strong></td>
</tr>
<tr>
<td>- Changes in the environment, human demographics, globalization and immunosuppressive therapies allow for the introduction of new viruses into the human population.</td>
</tr>
<tr>
<td>- Many chronic diseases may have an unrecognized viral etiology.</td>
</tr>
<tr>
<td>- Many human viruses remain undiscovered.</td>
</tr>
<tr>
<td><strong>Traditional strategies</strong></td>
</tr>
<tr>
<td>- Electron microscopy, virus culture and serology have been used to discover many viruses, even in modern times.</td>
</tr>
<tr>
<td>- However, these methods are often laborious and lack analytical sensitivity in comparison with modern molecular detection techniques, such as nucleic acid amplification tests.</td>
</tr>
<tr>
<td><strong>Viral nucleic acid extraction</strong></td>
</tr>
<tr>
<td>- Efficient concentration and recovery of viral nucleic acids from clinical specimens is the most crucial step in all molecular detection assays.</td>
</tr>
<tr>
<td><strong>Nucleic acid amplification tests</strong></td>
</tr>
<tr>
<td>- New members of the known virus families have been discovered through PCR tests targeting conserved regions in their genomes.</td>
</tr>
<tr>
<td>- The algorithmic approach to virus discovery involves the use of PCR to rule out the most common viral etiologic agents, followed sequentially by pan-viral microarray analysis and high-throughput sequencing.</td>
</tr>
<tr>
<td><strong>Differential display &amp; subtraction techniques</strong></td>
</tr>
<tr>
<td>- These methods required two sets of related samples, one potentially infected by the virus, and one that is uninfected.</td>
</tr>
<tr>
<td>- However, they can be more laborious than the use of microarrays and lack analytical sensitivity relative to high-throughput sequencing.</td>
</tr>
<tr>
<td><strong>Pan-viral microarrays</strong></td>
</tr>
<tr>
<td>- Pan-viral microarrays typically use long oligonucleotides that are more tolerant of mismatches and, thus, more suited to detecting novel viruses.</td>
</tr>
<tr>
<td>- Pan-viral microarrays employ a large number of probes, representing all known plant, animal and human viruses.</td>
</tr>
<tr>
<td>- Probes are designed to target both conserved genetic elements and species-specific regions in viral genomes.</td>
</tr>
<tr>
<td>- Computational algorithms are required to interpret the hybridization patterns on these arrays.</td>
</tr>
<tr>
<td><strong>Sequence-independent amplification</strong></td>
</tr>
<tr>
<td>- Sequence-independent amplification methods allow for the detection of novel viruses that are too divergent to be detected by other methods.</td>
</tr>
<tr>
<td>- These methods use random priming or the ligation of adaptor sequences to the sample DNA (or cDNA) in order to amplify the nucleic acids in an unbiased manner.</td>
</tr>
<tr>
<td>- The resulting library is then sequenced to determine the presence of new viruses.</td>
</tr>
<tr>
<td><strong>High-throughput sequencing</strong></td>
</tr>
<tr>
<td>- High-throughput sequencing is a sophisticated, yet ‘brute force’ method to detect novel viral sequences in clinical samples.</td>
</tr>
<tr>
<td>- Hundreds of thousands to millions of sequences must be computationally analyzed to determine the presence of novel viruses, requiring significant computational processing power and storage.</td>
</tr>
<tr>
<td>- These methods are still relatively costly and, thus, are only used further downstream in the algorithmic approach to virus discovery.</td>
</tr>
<tr>
<td><strong>Evidence of causation</strong></td>
</tr>
<tr>
<td>- Fulfillment of Koch’s postulates is often difficult for novel viruses and culture-independent methods are necessary to provide evidence of causality.</td>
</tr>
<tr>
<td>- However, the amplification of the virus, either through culture or generation of an infectious clone, remains the most essential tool for proving disease causation.</td>
</tr>
<tr>
<td><strong>Future perspective</strong></td>
</tr>
<tr>
<td>- Systematic approaches to virus discovery will continue to evolve as rapid advances in molecular detection platforms and metagenomics techniques continue.</td>
</tr>
<tr>
<td>- These methods will also allow us to characterize and better understand the human virome and its role in human health and disease.</td>
</tr>
</tbody>
</table>
epidemiological criteria, such as whether the nucleic acid sequence corresponding to the virus is present in most cases of an infectious disease, or whether virus genome copy number correlates with severity of disease or pathology [64,139,140]. For ubiquitous viruses (e.g., Epstein–Barr virus, human herpesvirus 6 and torque teno virus) or diseases in which host genetics and environmental factors play a significant role in the development of disease, proof of viral causation may be exceedingly difficult to obtain [141–143].

The ability to amplify a novel virus, either through cell culture directly from clinical samples or construction of an infectious clone, can be helpful not only in assessing causality but also in the detailed characterization of the virus. Cell culture experiments can be performed to assess host range and tissue and cell tropism. The mechanism of pathogenesis can be ascertained through studying the host–virus interaction at the molecular level. Concentrated titers of the virus can be used in ultrastructural studies or to produce antigens to develop assays for determining seroprevalence of the novel virus, either via the classical method of virus neutralization or via enzyme immunoassays. By providing a wealth of molecular and epidemiological data, these culture-dependent studies typically produce the most unambiguous evidence supporting the etiologic role of the novel virus in the infectious disease [144].

**Future perspective**

As the technology for metagenomic analysis continues to advance, so will our approach to virus detection and discovery. Automation and miniaturization of nucleic acid amplification and detection platforms, incorporation of microfluidics, and development of more sensitive hybridization methods all hold the promise of moving advanced molecular technology into the realm of point-of-care testing by clinicians [145–147]. The rapidly evolving techniques of next-generation sequencing will soon provide more affordable, higher throughput, single-molecule sequencing, such that sequence-based detection and diagnosis may eventually supplant nucleic acid amplification and microarray hybridization techniques [148]. Along with advances in analytical techniques, new methods for preanalytical sample treatment, such as dielectrophoresis for the separation of viral particles, and more efficient methods to concentrate nucleic acids from samples will be crucial tools to improve the rates of virus detection and discovery [149,150]. Our enhanced ability to identify and characterize more viruses, many of which still remain unknown, will allow us to appreciate the full extent of their interactions with humans and other living organisms [146,151].

Akin to efforts to characterize the ocean viromes and the human microbiome, similar virus surveys must also be done to determine the human virome [29]. Although not formally represented in the tree of life, viruses should be considered an essential part of the human metagenome, analogous to our growing recognition of the role of the human microbiome in health and disease [53].

**Financial & competing interests disclosure**

Charles Chiu is the Director of the UCSF-Abbott Viral Diagnostics and Discovery Centre. The Centre is funded by a research award from Abbott Diagnostics, Inc. (IL, USA). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

**Bibliography**

Papers of special note have been highlighted as:

- of interest
- of considerable interest


Review

Tang & Chiu


22. Detailed review of the role of electron microscopy in virus detection.


27. Comprehensive review of emerging and re-emerging pathogens and the role of molecular diagnostics.


32. Comparative review of new infectious agents and their role in disease of unknown etiology.


36. Outlines the need for active surveillance against new zoonotic diseases.


**Comprehensive review of the human virome and its role in human disease.**


**Comprehensive review of sequencing-based methods for virus discovery.**


**Description of contemporary methods to process samples for viral metagenomic studies.**


99. Comparison of differential display techniques.


Review of second- and third-generation DNA sequencing platforms.


Demonstration of single-molecule DNA sequencing of a virus.


Affiliations

Patrick Tang
British Columbia Centre for Disease Control, Department of Pathology & Laboratory Medicine, University of British Columbia, 655 West 12th Avenue, Vancouver, BC, V5Z 4R4, Canada
Tel.: +1 604 707 2616
Fax: +1 604 707 2675
patrick.tang@bccdc.ca

Charles Chiu
Departments of Laboratory Medicine & Medicine (Infectious Diseases), University of California, San Francisco, 185 Berry Street, Box 0134, San Francisco, CA 94143-0134, USA
Tel.: +1 415 514 8219
Fax: +1 415 514 8242
charles.chiu@ucsf.edu