Journal of Clinical Virology 61 (2014) 9–19

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

Review

Deep sequencing: Becoming a critical tool in clinical virology

Miguel E. Quiñones-Mateu [a,b,*], Santiago Avila [c,d], Gustavo Reyes-Teran [c,d], Miguel A. Martinez [e]

[a] University Hospital Translational Laboratory, University Hospitals Case Medical Center, Cleveland, OH, USA
[b] Department of Pathology, Case Western Reserve University, Cleveland, OH, USA
[c] Instituto Nacional de Enfermedades Respiratorias, Mexico City, Mexico
[d] Centro de Investigaciones en Enfermedades Infecciosas, Mexico City, Mexico
[e] Fundació irsicaixa, Universitat Autònoma de Barcelona, Hospital Universitari Germans Trias i Pujol, Badalona, Spain

ARTICLE INFO

Article history:
Received 9 April 2014
Received in revised form 12 June 2014
Accepted 14 June 2014

Keywords:
Next generation sequencing (NGS)
Deep sequencing
Human immunodeficiency virus (HIV)
Hepatitis C virus (HCV)
Influenza virus

ABSTRACT

Population (Sanger) sequencing has been the standard method in basic and clinical DNA sequencing for almost 40 years; however, next-generation (deep) sequencing methodologies are now revolutionizing the field of genomics, and clinical virology is no exception. Deep sequencing is highly efficient, producing an enormous amount of information at low cost in a relatively short period of time. High-throughput sequencing techniques have enabled significant contributions to multiples areas in virology, including virus discovery and metagenomics (viromes), molecular epidemiology, pathogenesis, and studies of how viruses to escape the host immune system and antiviral pressures. In addition, new and more affordable deep sequencing-based assays are now being implemented in clinical laboratories. Here, we review the use of the current deep sequencing platforms in virology, focusing on three of the most studied viruses: human immunodeficiency virus (HIV), hepatitis C virus (HCV), and influenza virus.

Contents

1. Introduction . . . 9
2. Current deep sequencing platforms . . . 10
3. Applications in general virology . . . 10
   3.1. Virus discovery . . . 10
   3.2. Human, animal, plant, and environmental viromes . . . 10
   3.3. Molecular epidemiology . . . 11
   3.4. Viral diversity, transmission, and pathogenesis . . . 12
4. Deep sequencing in the clinical setting . . . 13
   4.1. Drug resistance . . . 13
   4.2. HIV tropism . . . 14
   4.3. Implications of deep sequencing in clinical management . . . 14
5. Conclusions . . . 14
   Funding . . . 14
   Competing interests . . . 15
   Ethical approval . . . 15
   Acknowledgements . . . 15
   References . . . 15

1. Introduction

DNA sequencing has evolved considerably during the last 50 years. The first sequence of a tRNA molecule in 1965 [1] led to the

* Corresponding author at: Case Western Reserve University, Pathology, 10900 Euclid Avenue, Cleveland, OH 44106-7288, USA. Tel.: +1 216 286 4283; fax: +1 216 201 5432.

http://dx.doi.org/10.1016/j.jcv.2014.06.013
1386-6532/

development of more robust DNA sequencing methodologies in 1977 by Walter Gilbert and Allan Maxam (chemical degradation) [2] and Frederick Sanger (chain-termination) [3]. Sanger et al. determined the first complete genome sequence of any organism: the bacteriophage φ174 [3] which was followed by the DNA sequencing of a multitude of organisms, including DNA and RNA viruses, such as Epstein–Barr virus in 1984 [4] and Human Immunodeficiency Virus in 1985 [5]. Thirty years later, more than 2.5 million viral nucleotide sequences have been deposited in GenBank (https://www.ncbi.nlm.nih.gov/genbank/). This information has been instrumental to understand the structure of viral genomes, their biology, evolution, diversity, transmission, pathogenesis, as well evasion from the host immune response, antiviral drugs, and vaccines. Despite the impressive success of "Sanger" or "population" sequencing, and the development of automated DNA sequencing instruments in the mid 1980s, these first-generation sequencing methods are not high-throughput, have limited scalability and are not cost-effective for sequencing a multitude of samples and/or large genomes. A second generation of sequencing technologies has now addressed these limitations.

Following the development of several novel methods for DNA sequencing in the late 1990s, early 2000s [6,7], different technologies have become commercially available for so-called next generation sequencing (NGS), also known as "second generation", "massive parallel" or "deep" sequencing. Deep sequencing technologies are able to generate three–four orders of magnitude more information than Sanger sequencing and are significantly less expensive, considering the cost per nucleotide sequenced [8,9]. Several deep sequencing systems have been developed during the last 10 years, each of them with their own intrinsic performance metrics such as number of reads obtained, read length, accuracy, time to run, cost, etc. [9,10]. In 2004, 454 Life Sciences (Branford, CT) introduced the first instrument based on pyrosequencing [11], prior to being acquired by Roche (Basel, Switzerland). This was followed by the release of the Genome Analyzer (GA, Solexa, Chesterford, UK) in 2005 [12], now Illumina (San Diego, CA). In 2007, the first SOLiD sequencing system was released by Applied Biosystems (Foster City, CA) [13], while the Helicos sequencer [14] and the Ion Torrent Personal Genome Machine [15], were commercialized by Life Technologies (Carlsbad, CA) in 2009 and 2011, respectively. Pacific Biosciences (Menlo Park, CA) introduced the single molecule real-time sequencer in 2011 [16] and more recently Oxford Nanopore Technologies (Oxford, UK) released an ultra-long single molecule sequencer [17]. All these methodologies have continued to evolve, with new chemistries and more potent instruments, resulting in impressive levels of sequencing throughput at constantly lower costs. As a result, the use of deep sequencing continues to expand both in research and clinical settings, and the virology field is no exception [18–25]. In this review we provide an overview of the use of deep sequencing in virology and how these techniques have impacted, and will continue to influence, the study of three of the most important human viral agents that top the list of viral nucleotide sequences in GenBank, i.e., human immunodeficiency virus (HIV), hepatitis C virus (HCV), and influenza virus.

2. Current deep sequencing platforms

Many excellent articles have compared and described on detail the different deep sequencing methodologies and instruments, including methods used in template preparation, sequencing, and data analysis [9,22]. Four platforms dominate the deep sequencing field: 454™ (454 Life Sciences/Roche, Branford, CT) [11], Illumina® (Illumina, Inc., San Diego, CA) [12], Ion Torrent™ (Ion Torrent/Life Technologies, South San Francisco, CA) [15], and PacBio® (Pacific Biosciences, Menlo Park, CA)[16]. In general, all four technologies are able to generate valuable sequence information [26]; however, there are key and significant differences between the amount and quality of the data, and the applications that each system could support (Fig. 1). Selecting a deep sequencing platform depends on the potential application(s) and resources available, including cost of the instrument and reagents, existing infrastructure and personal experience. Although to date most of the published studies in virology have used 454™ or Illumina® systems, perhaps due to the fact that these were the first two methodologies available (Fig. 1), all deep sequencing technologies continue to improve and are being used in a multitude of virological studies [18–20,24].

3. Applications in general virology

Deep sequencing has been a "shot in the arm" to all genomic studies and virologists have taken particularly advantage of this methodology. First of all, now it is simpler and more affordable than ever to sequence full viral genomes. Likewise, identification and classification of novel and known viruses, unbiased characterization of viral populations without the need for virus culturing (viromes), molecular epidemiology, viral diversity and evolution, transmission and pathogenesis, and in particular medical virology have greatly benefited from the use of deep sequencing. As shown in Fig. 2A, the number of publications using deep sequencing in virology has skyrocketed since 2008, particularly those associated with HIV. By allowing the cost-effective study of a greater number of viral variants, including complex viral populations, deep sequencing is the perfect tool for a broad number of applications for studies of HIV, HCV and influenza virus (Fig. 2B).

3.1. Virus discovery

The field of virus discovery has flourished with the advent of deep sequencing methodologies. Coupled with bioinformatics tools, high throughput sequencing has revolutionized the field by allowing the identification and characterization of novel viruses [27] including a novel rhabdovirus associated with acute hemorrhagic fever identified in Central Africa [30] and a new cyclovirus detected in cerebrospinal fluid of patients with central nervous system infections [31]. Together with classical methods of diagnostics such as virus isolation, immunochemistry, and PCR, deep sequencing will continue to play a significant role in the identification of novel viruses, particularly in the face of outbreaks of known and/or new diseases.

3.2. Human, animal, plant, and environmental viromes

Viral metagenomic studies, i.e., the characterization of viral genomes directly from samples, have become extremely popular with the arrival of deep sequencing [28] and numerous viral metagenomic analyses have been published (reviewed in [18,27–29]). The viral metagenome or "virome" refers to the collection of viruses found in a particular sample from humans, animals, plants or from a specific environmental sample. Virome studies can lead to the discovery of new viruses and/or to their association with known or novel diseases. In the case of the human virome, its composition and impact on human health has been the subject of multiple studies [37,38]. The healthy human gut virome seems to consist mainly of bacteriophages [39], and can vary among individuals and in response to diet [40]. On the other hand, the human skin virome comprises a high variety of DNA viruses [41]. Interestingly, the human gut, nasopharyngeal, or plasma viromes are highly susceptible to changes associated with diseases such as HIV/AIDS [42,43], antiviral and immunosuppression therapy [44], or respiratory infections [45,46]. Similar studies have analyzed the