



Contents lists available at SciVerse ScienceDirect

Journal of Clinical Virology

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



Review

Next-generation sequencing technologies in diagnostic virology

Luisa Barzon*, Enrico Lavezzo, Giulia Costanzi, Elisa Franchin, Stefano Toppo, Giorgio Palù

Department of Molecular Medicine, University of Padova, Via A. Gabelli 63, I-35121 Padova, Italy

ARTICLE INFO

Article history:

Received 24 November 2012
Received in revised form 1 March 2013
Accepted 2 March 2013

Keywords:

Genome sequencing
Bioinformatics
Deep/high throughput sequencing
Virus discovery
Viral metagenomics/viromics
Viral epidemiology

ABSTRACT

The data deluge produced by next-generation sequencing (NGS) technologies is an appealing feature for clinical virologists that are involved in the diagnosis of emerging viral infections, molecular epidemiology of viral pathogens, drug-resistance testing, and also like to do some basic and clinical research. Indeed, NGS platforms are being implemented in many clinical and research laboratories, as the costs of these platforms are progressively decreasing. We provide here some suggestions for virologists who are planning to implement a NGS platform in their clinical laboratory and an overview on the potential applications of these technologies in diagnostic virology.

© 2013 Elsevier B.V. All rights reserved.

Contents

1. Introduction	00
2. Implementation of NGS technology in the clinical virology laboratory	00
3. Choice of the NGS platform	00
4. Pitfalls of NGS in diagnostic virology	00
5. Niches for NGS in diagnostic virology	00
6. Applications of NGS in diagnostic virology	00
6.1. Identification of novel or unsuspected viral pathogens	00
6.2. Metagenomics analysis of the human virome	00
6.3. Full-length viral genome sequencing	00
6.4. Ultra-deep targeted sequencing for the analysis of viral populations	00
6.5. Analysis of virus–host interaction	00
7. Bioinformatics	00
8. Challenges for clinical application of NGS technologies	00
9. Conclusions	00
Funding	00
Competing interests	00
Ethical approval	00
References	00

1. Introduction

Novel DNA sequencing techniques, referred to as “next-generation” sequencing (NGS), provide high speed and throughput that can produce an enormous volume of sequences in a single run at relatively low cost. The most important advantage of these platforms is the ability to determine the sequence data from single DNA fragments of a library, avoiding the need for cloning in vectors prior to sequence acquisition. These techniques have provided

Abbreviations: NGS, next generation sequencing; HIV, human immunodeficiency virus; HBV, hepatitis B virus; HCV, hepatitis C virus; EBV, Epstein-Barr virus; HCMV, human cytomegalovirus; HSV, herpes simplex virus.

* Corresponding author. Tel.: +39 0498218946; fax: +39 0498272355.
E-mail address: luisa.barzon@unipd.it (L. Barzon).

a great contribution to the research in many fields of life sciences¹ and are being increasingly introduced in clinical laboratories, with many diagnostic applications in human genetics, oncology, as well as in microbiology and virology.^{2,3} We provide here some suggestions for virologists who are planning to implement a NGS platform in their clinical laboratory and an overview on the potential applications of these technologies in diagnostic virology.

2. Implementation of NGS technology in the clinical virology laboratory

The data deluge produced by NGS technologies is an appealing feature for clinical virologists that are involved in the diagnosis of emerging viral infections, molecular epidemiology of viral pathogens, drug-resistance testing, and also like to do some basic and clinical research. Indeed, NGS platforms are being implemented in many clinical and research laboratories, as the costs of these platforms are progressively decreasing.

Different NGS platforms are available in the market, which are best suited for some applications but not for others. Thus, before acquiring a NGS platform, a careful analysis of the diagnostic and research needs of the laboratory should be done and expertise in NGS laboratory protocol set-up and NGS data assembly and analysis should be achieved. To this aim, we strongly recommend to establish a “wet lab team” and a “bioinformatics team” that collaborate to provide answers to the laboratory needs, before getting the NGS platform. Requirements of the “wet lab team” are skills in molecular biology techniques and biotechnologies, while requirements for the “bioinformatics team” are expertise in data management, implementation of sequence alignment algorithms, design of custom working pipelines and statistical analysis.

In the settlement of instruments and protocol workflow for NGS, care should be taken to separate pre- and post-PCR phases and to avoid the risk of nucleic acid contamination (e.g., during DNA nebulisation) also with the other conventional molecular diagnostic methods in the laboratory. Most important, in order to be used in routine testing, diagnostic tests performed by NGS technologies require analytical and clinical validation according to current guidelines and recommendations that are used for molecular assays and must be under the supervision of quality assurance and quality control programmes.

3. Choice of the NGS platform

Different NGS methods are commercially available and novel and improved platforms are continuously being developed and released. These NGS methods have different underlying biochemistries⁴ and differ in sequencing protocol, throughput, and sequence length.² Thus, the SOLiD system (Life Technologies), characterised by extremely high throughput but very short reads, may be more suitable for applications such as large whole genome resequencing or RNA-sequencing projects; while other platforms, like 454 (Roche Diagnostics), Ion Torrent (Life Technologies), and Illumina sequencing systems (Illumina) provide data suitable for de novo assembly, even though the relatively limited throughput of 454 and Ion Torrent PGM restricts their application to small bacterial- and viral-size genomes. In contrast, the relative long length of 454 FLX (and its smaller version GS Junior) reads allows deep sequencing of larger amplicons, with applications in microbial and viral metagenomics, analysis of viral quasispecies, and viral haplotypes reconstruction.⁵

Besides these platforms, which are now consolidated and spread in laboratories all around the world, new instruments have been recently launched (PacBio RS from Pacific Biosciences)⁶ or will be released in the near future (GridION and MinION from Oxford

Nanopore Technologies),^{7,8} the so called “third generation” technologies. These systems will probably take sequencing applications to a next level of performance, since they claim to be able to sequence single DNA molecules and reach read lengths above 10,000 bp, but a settling time will be needed before routine application in diagnostics.

The choice of the NGS platform for diagnostic virology should take into consideration the following aspects:

- *Diagnostic applications:* The types of diagnostic applications and the number of different tests to set up are relevant in the choice of a NGS platform. For a single application (e.g., deep sequencing of amplicons), small and relatively inexpensive instruments, like GS Junior and Ion Torrent PGM, are available. If the type and number of tests are not defined a priori and a flexible platform is required, Illumina sequencing systems provide several advantages, especially in terms of throughput, that could respond to most diagnostic needs, though reads are shorter than those from 454 FLX.
- *Costs:* The costs of tests by using NGS methods could be much higher or lower than those of conventional molecular methods and this should be taken into account. E.g., resequencing a full viral genome by conventional cycle sequencing is generally less expensive than by using NGS methods. Likewise, detection of a known single nucleotide mutation by real-time PCR is less expensive than by using deep sequencing of amplicons.
- *Speed:* The turnaround time of NGS protocols and data analysis is a critical issue if the NGS platform is to be used in “real-time” diagnostics, aiming to provide useful information for disease prevention or therapeutic interventions. Thus, for most diagnostic applications, the laboratory staff and/or the vendor should develop simplified laboratory protocols for library preparation and bioinformatics tools for easy data analysis and interpretation.
- *Throughput:* Different applications require different throughputs (e.g., shotgun metagenomics applications for pathogen detection require a much higher throughput than targeted resequencing applications).
- *Accuracy:* While a very high accuracy is needed for mutation detection (e.g., in drug resistance testing), this issue is less relevant for pathogen detection by shotgun approaches.
- *Read length:* Long reads are needed for deep sequencing of amplicons and de novo sequencing of genomes, and are also useful for pathogen discovery.
- *Upgrading:* NGS technologies are rapidly improving and evolving, thus upgradable instruments should be preferred.
- *Automation:* Since NGS protocol workflow is generally very long and complex, automation of library preparation protocols is strongly recommended.

4. Pitfalls of NGS in diagnostic virology

NGS technologies offer great chances for diagnostic virology, but there are several pitfalls in comparison with applications in diagnostic microbiology and other diagnostic applications.

- *Low sensitivity:* The sensitivity of NGS methods for the detection of viral sequences is relatively low due to the small size of the viral genome in comparison with host genome, that invariably contaminates clinical specimens or viral cultures. This is particularly relevant when viral load is very low. Thus, for viral genome detection, conventional PCR and cycle sequencing may be more convenient than NGS sequencing.
- *Whole viral genome sequencing:* The presence of contaminating host genome sequences is also a problem for whole viral genome sequencing. To enrich viral particles in the sample, contaminating

genomes may be removed by nucleic acid digestion, ultracentrifugation, and filtration. Viral genome sequences may be also directly enriched by PCR amplification with primers targeting conserved sequences, by rolling cycle amplification, or by target sequence capture.^{2,9} In all cases, a post-processing analysis step is required to filter host-related reads and contaminants.¹⁰

- *Viral genome variability*: Targeted resequencing can be implemented with NGS protocols, if the target virus is known. However, like with conventional PCR and sequencing methods, the high variability of viral genomes represents a problem for the design of targeted resequencing protocols to be used in diagnostic virology. In addition, at variance with bacteria and eukaryotes, for which very conserved target sequences are available that allow population and metagenomics studies (e.g., genes encoding ribosomal RNA), this is not feasible for viruses, for which conserved sequences can only be designed at genus or family level.

5. Niches for NGS in diagnostic virology

NGS will not replace conventional virology, and viral isolation in cell culture, viral detection by molecular testing, and genetic analysis by cycle sequencing will probably continue to be standard diagnostic methods for the next years. Instead, NGS methods are expected to be complementary to conventional diagnostic methods and to be used for niche applications, providing a powerful tool to face the challenges of old and emerging viral infections. These include:

- *Pathogen discovery* by metagenomics approaches;
- *Analysis of intra- and inter-host virus variability* (e.g., drug-resistance testing, virus genotyping, molecular epidemiology of viral infections, anti-viral drug and vaccine discovery);
- *Virus–host interaction studies* (e.g., transcriptome analysis by RNA sequencing, epigenomics analyses, studies of protein–nucleic acid interactions);
- *Host genetics and infectious disease susceptibility* (e.g., investigation of genetic traits associated with the risk of viral diseases).

6. Applications of NGS in diagnostic virology

6.1. Identification of novel or unsuspected viral pathogens

The most interesting and fascinating application of NGS technologies in diagnostic virology is probably the identification of novel or unsuspected viral pathogens. Examples of this application include:

- Real-time investigation of outbreaks (e.g., the discovery of Schmallenberg virus, a novel orthobunyavirus responsible for congenital malformations in cattle and sheep^{11,12});
- Etiologic diagnosis of viral infections by using a metagenomic approach (e.g., the recent identification of viruses responsible for unknown severe febrile diseases^{13,14});
- Discovery of new tumor viruses (e.g., Merkel Cell polyomavirus, the etiologic agent of Merkel cell carcinoma¹⁵);
- Discovery of new potentially pathogenic viruses (e.g., discovery of a new filovirus in bats¹⁶).

With the application of shotgun sequencing methods for metagenomics studies, new viruses, including non-pathogenic commensal viruses, are being discovered. Thus, experimental and epidemiological verification of the causative role of the newly discovered virus is always required.

6.2. Metagenomics analysis of the human virome

The composition of the human virome in health and disease is being deciphered by the use of NGS methods.^{17,18} But, unlike the human microbiome, the human virome is still largely unexplored. In this field, NGS will provide a great contribution to our understanding of the complex interplay between viruses and their human host.

6.3. Full-length viral genome sequencing

Diagnostic application of full-length viral genome sequencing include the following investigations:

- Viral genome diversity and evolution (e.g., investigation of HIV evolution and spread¹⁹);
- Viral phylogenesis (e.g., influenza virus²⁰);
- Molecular epidemiology and surveillance (e.g., arboviruses and respiratory viruses²¹);
- Pathogenic mutations and fitness (e.g., identification of mutations in HIV and HCV genome that affect viral fitness and outcome of antiviral therapies^{22,23});
- Variants escaping host neutralization (e.g., HIV and other viruses²⁴; vaccine efficacy monitoring)
- Intra-host viral genome diversity (e.g., viruses responsible of chronic infections, like HIV, HBV, HCV, herpesviruses^{25,26});
- Identification of virulence factors and drug-resistance genes (e.g., HSV, HCV, HIV, HBV, HCMV).^{24–27}

6.4. Ultra-deep targeted sequencing for the analysis of viral populations

- Virus genotyping (e.g., human papillomaviruses²⁸)
- Identification of quasispecies (e.g., characterization of HCV and HIV quasispecies; reviewed in Ref. [2]);
- Detection of rare drug-resistant variants (e.g., drug-resistance testing for HIV, HCV, and HBV; reviewed in Ref. [2]).

6.5. Analysis of virus–host interaction

Analysis of virus–host interaction is generally a field for basic science investigation. However, improvements in the knowledge on the mechanism of viral disease could lead to the identification of useful diagnostic and prognostic markers for viral infections, identification of potential targets for antiviral drugs, and design of new vaccines. The contribution of NGS to the analysis of virus–host interaction and pathogenesis includes:

- RNA-sequencing analysis of viral gene expression (e.g., identification of new viral genes and transcript isoforms during EBV and HCMV infection using RNA-sequencing^{29,30});
- Viral and host transcriptome and epigenome analysis (e.g., characterization of host-virus regulatory interactions during EBV infection³¹).

7. Bioinformatics

A common feature of NGS technologies is the extremely high throughput data generation. This, however, was obtained at the expense of the accuracy, quality, and average length of the produced sequences, that are much lower than that obtained with the Sanger chemistry, leading inevitably to overcome several hurdles. As a result, new issues have to be addressed in order to exploit the full potential of these new instruments: firstly, the data analysis step has become very time consuming and novel approaches

have been designed to manage short reads, requiring a competent amount of manpower and expertise in bioinformatics; secondly, adequate computing resources are necessary to handle the data produced, especially when they are applied in diagnostic testing.

The role of bioinformatics is pivotal, because it merges two distant realities, biology and informatics. Bioinformaticians must be able to bridge the differences between experts and make right communication happen. Bioinformatics expertise is generally acquired after graduation; training courses or masters are presently organized in many universities and institutions. People teaching in such courses generally come from both biology and informatics, and also students should possess different academic backgrounds; this mixing is fruitful in terms of knowledge exchange and reflects the actual situation that it would be desirable to achieve in the “bioinformatics team”.

8. Challenges for clinical application of NGS technologies

Indeed, NGS data management requires to address very different tasks, sometimes strictly computational, sometimes more related to biology, and a team must possess the expertise in both fields. This is particularly relevant in diagnostic virology and there is an obvious need for joint academic and clinical diagnostic partnerships to fully enable this promising technology, similar to how clinical virology departments originally emerged from an Academic environment. The more demanding regulatory environment that we now live in concerning governance, ethics, corporate goals, could pose real stumbling blocks to this type of partnerships. Efforts should be done by both Academic and Clinical experts to collaborate in the development of tools to face the new challenges posed by NGS technologies in the clinical practice. Clinical data generated by NGS technology should not be considered qualitatively different from any other clinical test, since both clinical NGS and conventional tests are used for making a diagnosis or a prognosis. However, clinical data provided by NGS, even if applied to diagnostic virology, might incidentally led to the discovery of unsuspected or unwanted genetic data that will require interpretation. Conceptually, the same problem, although at a smaller extent, is posed by any other open-ended conventional clinical test, like karyotype analysis, imaging procedures, or general physical examination, which have the risk of serendipitous discovery of new clinical problems and, as a cascade effect, requests of new diagnostic tests. To avoid this risk, clinical NGS tests should be set up as closed tests targeted to a particular diagnostic problem. In addition, clinical NGS data should be interpreted and delivered to patients by expert professionals that are familiar with test methodology, bioinformatics analysis of data, and potential pitfalls related to the technology. Because of all these problems and, in addition, issues related to data security and privacy, patients should receive adequate information and counselling before NGS testing.³²

9. Conclusions

NGS technologies have many potential applications in diagnostic virology, such as discovery and characterization of new viruses, detection of unexpected viral pathogens in clinical specimens, ultrasensitive monitoring of antiviral drug resistance, investigation of viral diversity, evolution and spread, evaluation of the human virome, and investigation on virus–host interaction and pathogenesis of viral diseases. With the decrease of costs and improvement of turnaround time, these techniques will probably become essential diagnostic tools in clinical routines. Standardization of NGS methods, together with analytical and clinical validation of tests, are key issue for the implementation of NGS technologies in the clinical laboratory.

Funding

This study was supported by the European Commission under FP7, Project 261426 (WINGS West Nile Integrated Shield Project).

Competing interests

None declared.

Ethical approval

Not required.

References

1. Kahvejian A, Quackenbush J, Thompson JF. What would you do if you could sequence everything? *Nat Biotechnol* 2008;**26**:1125–33.
2. Barzon L, Lavezzo E, Militello V, Toppo S, Palù G. Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci* 2011;**12**:7861–84.
3. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. *J Gen Virol* 2012;**93**:1853–68.
4. Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet* 2010;**11**:31–46.
5. Zagordi O, Däumer M, Beisel C, Beerenwinkel N. Read length versus depth of coverage for viral quasispecies reconstruction. *PLoS ONE* 2012;**7**:e47046.
6. Eid J, Fehr A, Luong K, Lyle J, Otto G, Peluso P, et al. Real-time DNA sequencing from single polymerase molecules. *Science* 2009;**323**:133–8.
7. Eisenstein M. Oxford Nanopore announcement sets sequencing sector abuzz. *Nat Biotechnol* 2012;**30**:295–6.
8. Schneider GF, Dekker C. DNA sequencing with nanopores. *Nat Biotechnol* 2012;**30**:326–8.
9. Bexfield N, Kellam P. Metagenomics and the molecular identification of novel viruses. *Vet J* 2011;**190**:191–8.
10. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RG, Getz G, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol* 2011;**29**:393–6.
11. Hoffmann B, Scheuch M, Höper D, Jungblut R, Holsteg M, Schirrmeier H, et al. Novel orthobunyavirus in Cattle, Europe, 2011. *Emerg Infect Dis* 2012;**18**:469–72.
12. Rosseel T, Scheuch M, Höper D, De Regge N, Caij AB, Vandenbussche F, et al. DNase SISPA-next generation sequencing confirms Schmallenberg virus in Belgian field samples and identifies genetic variation in Europe. *PLoS ONE* 2012;**7**:e41967.
13. Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E, DeRisi JL. Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis* 2012;**6**:e1485.
14. McMullan LK, Folk SM, Kelly AJ, MacNeil A, Goldsmith CS, Metcalfe MG, et al. A new phlebovirus associated with severe febrile illness in Missouri. *N Engl J Med* 2012;**367**:834–41.
15. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008;**319**:1096–100.
16. Negro A, Palacios G, Vázquez-Morón S, González F, Dopazo H, Molero F, et al. Discovery of an ebolavirus-like filovirus in Europe. *PLoS Pathog* 2011;**7**:e1002304.
17. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 2011;**21**:1616–25.
18. Reyes A, Semenkovich NP, Whiteson K, Rohwer F, Gordon JI. Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat Rev Microbiol* 2012;**10**:607–17.
19. Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis* 2011;**11**:45–56.
20. Kuroda M, Katano H, Nakajima N, Tobiume M, Aina I, Sekizuka T, et al. Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PLoS ONE* 2010;**5**:e10256.
21. Baillie GJ, Galiano M, Agapow PM, Myers R, Chiam R, Gall A, et al. Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. *J Virol* 2012;**86**:11–8.
22. Bull RA, Luciani F, McElroy K, Gaudieri S, Pham ST, Chopra A, et al. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathog* 2011;**7**:e1002243.
23. Ninomiya M, Ueno Y, Funayama R, Nagashima T, Nishida Y, Kondo Y, et al. Use of Illumina deep sequencing technology to differentiate hepatitis C virus variants. *J Clin Microbiol* 2012;**50**:857–66.
24. Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, et al. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS ONE* 2010;**5**:e12303.

25. Gorzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E. Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. *J Virol* 2010;**84**:7195–203.
26. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. *PLoS Pathog* 2011;**7**:e1001344.
27. Szpara ML, Parsons L, Enquist LW. Sequence variability in clinical and laboratory isolates of herpes simplex virus 1 reveals new mutations. *J Virol* 2010;**84**:5303–13.
28. Barzon L, Militello V, Lavezzo E, Franchin E, Peta E, Squarzon L, et al. Human papillomavirus genotyping by 454 next generation sequencing technology. *J Clin Virol* 2011;**52**:93–7.
29. Concha M, Wang X, Cao S, Baddoo M, Fewell C, Lin Z, et al. Identification of new viral genes and transcript isoforms during Epstein-Barr virus reactivation using RNA-Seq. *J Virol* 2012;**86**:1458–67.
30. Gatherer D, Seirafian S, Cunningham C, Holton M, Dargan DJ, Baluchova K, et al. High-resolution human cytomegalovirus transcriptome. *Proc Natl Acad Sci USA* 2011;**108**:19755–60.
31. Arvey A, Tempera I, Tsai K, Chen HS, Tikhmyanova N, Klichinsky M, et al. An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell Host Microbe* 2012;**12**:233–45.
32. Biesecker LG, Burke W, Kohane I, Plon SE, Zimmern R. Next-generation sequencing in the clinic: are we ready? *Nat Rev Genet* 2012;**13**:818–24.