HAYATI Journal of Biosciences 23 (2016) 155–159

HOSTED BY

Contents lists available at ScienceDirect

HAYATI Journal of Biosciences

journal homepage: http://www.journals.elsevier.com/hayati-journal-of-biosciences

Review paper

Next-Generation Sequencing and Influenza Virus: A Short Review of the Published Implementation Attempts

Rasha Ali,[1] Ruth Marion Blackburn,[2] Zisis Kozlakidis[1,2*]

[1] Division of Infection and Immunity, University College London, London, United Kingdom.
[2] Farr Institute of Health Informatics Research, University College London, London, United Kingdom.

ARTICLE INFO

Article history:
Received 23 May 2016
Accepted 9 December 2016
Available online 11 January 2017

KEYWORDS:
implementation,
influenza,
next generation sequencing

ABSTRACT

Influenza virus represents a major public health concern worldwide after recent pandemics. To aid the understanding and characterization of the virus in ever-increasing sample numbers, new research techniques have been used, such as next-generation sequencing (NGS). The current article review used Ovid MEDLINE and PubMed databases to conduct keyword searches and investigate the extent to which published NGS high-throughput approaches have been implemented to influenza virus research in the last 5 years, during which the increase in research funding for influenza studies has been coincidental with a significant per-base cost reduction of sequencing. Through the current literature review, it is evident that over the last 5 years, NGS techniques have been indeed applied to biological and clinical samples at increasing rates following a wide variety of approaches. The rate of adoption is slower than anticipated by most published studies, with three obstacles identified consistently by authors. These are the lack of suitable downstream analytical capacity, the absence of established quality control comparators, and the higher cost to comparable existing techniques.

1. Introduction

Influenza viruses are well-characterized members of the Orthomyxoviridae family. Genomic subpopulation diversity and new viral mutants emerge constantly because of the continued viral genetic variation and antigenic modification in response to many factors such as host immunity, ecological and environmental factors, resulting in occasional pandemics and annual epidemics (Zhirnov *et al.* 2009). Influenza remains a major threat on the global agricultural and health care systems because of its continued potential to cause pandemics worldwide and because of the increasing number of seasonal infections impacting human and economic health (Fischer *et al.* 2015). The high number of infections and the recurrent seasonality mean that influenza is suitable for a number of high-throughput molecular approaches in addition to the basic virological techniques and clinical expertise to strengthen global pandemic preparedness. In addition, the total and proportional funding for influenza research (£39,139,703, 4.3% of total infection research) increased in 2011–13 compared with 1997–2010 (£126,643,152, 3.4% of all infection research), hence the field is more likely able to afford the use of new and perhaps more expensive technologies than studies of other infectious diseases (Head *et al.* 2016). Coincidentally, the per-base cost of sequencing in the same period has reduced by 92% from 0.52 to 0.04 USD per DNA Mb (National Human Genome Research Institute, January 2010–January 2015). Hence, according to our working hypothesis, we expected to notice a steady increase in published implementation examples as overall implementation costs were reducing. In this brief report, we review the application of high-throughput next-generation sequencing (NGS) in the study of influenza and present the opportunities and challenges of implementation as reported by the research community.

Currently, there are two major technologies used for influenza genomic sequencing; the NGS and traditional Sanger sequencing (Deng *et al.* 2015). The Sanger sequencing technology referred to as first generation has been used for almost four decades and continues to be the standard reference method used. However, there is a gradual yet notable shift away from this technique and in favor of the use of newer technologies, namely the high-throughput NGS (International Human Genome Consortium 2004). NGS also referred to as deep sequencing or parallel sequencing (massively parallel sequencing) provides high-speed multiplexing capabilities

* Corresponding author.
E-mail address: z.kozlakidis@ucl.ac.uk (Z. Kozlakidis).
Peer review under responsibility of Institut Pertanian Bogor.

http://dx.doi.org/10.1016/j.hjb.2016.12.007

for high-throughput sample sequencing and enormous data volumes of sequencing reads in one run (Barzon *et al.* 2011). Along with the decreasing NGS costs, the applications of NGS techniques within routine diagnostic settings are still evolving because of recent and iterative developments in genome sequencing and bioinformatics analyses (Fischer *et al.* 2015).

2. A Number of choices and challenges for NGS platforms

The common process of most NGS technologies is the initial random fragmentation of templates, followed by an amplification process using polymerase chain reaction target-specific primers, resulting in many DNA copies that can be independently sequenced (Metzker 2010). High-throughput sequencing platforms can be divided into two broad groups depending on the template used. The earliest platforms depend on the production of libraries of clonally amplified templates. The recent arrival of single-molecule sequencing platforms determines the sequence of single molecules without amplification. Within these broad categories, there is considerable variation in performance—including in throughput, read length, and error rate—as well as in factors affecting usability, such as cost and run time (Loman *et al.* 2012).

NGS technologies have a unique potential for the *de novo* sequencing of large genomes, genomic markers screening, transcriptome analysis, and several other applications (Bainbridge *et al.* 2006; Cheval *et al.* 2011; Greninger *et al.* 2010; Kuroda *et al.* 2010; Nakamura *et al.* 2009; Pettersson *et al.* 2008; Satkoski *et al.* 2008; Torres *et al.* 2008; Wheeler *et al.* 2008). However, the complexity and large size of the sequencing data constitute one of the main bioinformatics challenges of NGS data interpretation (Nowrousian 2010). The primary approach to NGS data analysis can be accomplished by using either one of three main types of tools, such as general-purpose aligners, *de novo* assemblers, and short-read aligners (Lin *et al.* 2014). NGS methods confer advantages over other techniques such as highly specific reverse transcription-polymerase chain reaction or less-sensitive traditional virological methods for being able to produce unbiased sequencing without prior knowledge of the presence or type of viral agents. This in turn can potentially constitute them into the future gold standard tool for viral genome discovery, especially in the case of recombinogenic viruses, such as influenza (Bialasiewicz *et al.* 2014).

Through the current literature review, it is evident that over the last 5 years, NGS techniques have been indeed applied to clinical samples at increasing rates with some studies concentrating on the detection of novel pathogens or pathogens at low detection levels. Several variant strains and viruses have been successfully identified, such as the PIV4 subtype in late 2013(Alquezar-Planas *et al.* 2013), although it has to be noted that the numbers of unsuccessful attempts are generally not mentioned, unclear, and/or very difficult to even hazard a guess at. Other studies followed the seasonal influenza infections in large population cohorts (Nakamura *et al.* 2009), whereas influenza studies on animals have also used NGS capabilities, such as sequences generated from lung tissues of ferrets experimentally infected with influenza A/California/07/2009 (H1N1) (Lin *et al.* 2014). However, the overall numbers of samples used per study vary widely, and the full implementation of a high-throughput analytical pipeline remains difficult to achieve. The implementation challenges, solutions, and expectations of the authors are also summarized.

3. Methods

Our research based on the Ovid MEDLINE database and the NCBI PubMed databases was conducted with a total of 18 different keywords in different combinations each time (initial concept terms used: Influenza, next generation sequencing, and data not shown). The literature search provided a wide variety of peer-reviewed publications ranging in number from (10−18013). The relevant article abstracts were manually selected corresponding to publications where NGS was actually implemented as opposed to being alluded to for future implementation. Then the exact sequencing techniques used were determined, e.g. Illumina™ MiSeq/HiSeq NGS, Roche™ GS-FLX+ 454-pyrosequencing, and others. Only two inclusion criteria were preselected, that is English language and publication years from 2008 to 2015 inclusive.

4. Results

4.1. Influenza high-throughput DNA sequencing studies

Our research detected 64 research publications within the publication years of 2008−2015. According to their methods, almost all the studies used one or more of the following NGS platforms (Roche-454 GS Junior/FLX+, Ion Torrent/Proton/Personal Genome Machine sequencing, and Illumina GAIIx/MiSeq/HISeq) accompanied with a diverse and fragmented set of methods for the upstream sample preparation and downstream bioinformatics analyses.

Of the 64 research publications, 35 studies were performed exclusively on human material (Fischer *et al.* 2015; Deng *et al.* 2015; Kuroda *et al.* 2010; Cheval *et al.* 2011; Buggele *et al.* 2013; Depew *et al.* 2013; Baum *et al.* 2010; Rutvisuttinunt *et al.* 2015; Frey *et al.* 2014; Farsani *et al.* 2015; Zhao *et al.* 2015; Rutvisuttinunt *et al.* 2013; Lee *et al.* 2013; Flaherty *et al.* 2012; Téllez-Sosa *et al.* 2013; Borozan *et al.* 2013; Archer *et al.* 2012; Bidzhieva *et al.* 2014; Van den Hoecke *et al.* 2015; Leung *et al.* 2013; Watson *et al.* 2013; Harismendy *et al.* 2009; Zhou *et al.* 2014; Kuroda *et al.* 2015; Burnham *et al.* 2015; Varble *et al.* 2014; Tan *et al.* 2014; Saira *et al.* 2013; Selleri, 2013; Swaminathan *et al.* 2013; Xiao *et al.* 2013; Power *et al.* 2012; Whitehead *et al.* 2012; Yasugi *et al.* 2012), 10 on animal material (Lin *et al.* 2014; Jakhesara *et al.* 2014; Van Borm *et al.* 2012; Dugan *et al.* 2011; Clavijo *et al.* 2013; León *et al.* 2013; Lange *et al.* 2013; Iqbal *et al.* 2014; Peng *et al.* 2011; Wang *et al.* 2012), seven on both animal and human materials (Yu *et al.* 2014; Jonges *et al.* 2014; Kampmann *et al.* 2011; Peng *et al.* 2014; Karlsson *et al.* 2013; Sikora *et al.* 2014; Ren *et al.* 2013), two on plasmid-derived material (Depew *et al.* 2013; Wu *et al.* 2014), and 10 reviewed technical and bioinformatics aspects (Barzon *et al.* 2011; Metzker 2010; Quiñones-Mateu *et al.* 2014; Park *et al.* 2013; Dugan *et al.* 2012; MacLean *et al.* 2009; Radford *et al.* 2012; Ansorge 2009; Shendure and Ji 2008; Tsai and Chen 2011). The number of samples used per study varied widely, with most studies reporting numbers in the low hundreds and less than 10 reporting the use of more than 1000 samples.

4.2. Challenges, opportunities, and solutions of NGS implementation

From the aforementioned, it becomes immediately obvious that the initial NGS applications in the field of influenza research are not reflective of a consistent, universally applied, and true high-throughput approach. Indeed, the picture obtained throughout is one reflecting the initial stages for the adoption of a technical innovation. The challenges mentioned by the various authors are summarized in the Table. The generation of high volumes of data requiring sophisticated downstream bioinformatics analyses is mentioned as the primary challenge for the adoption of the method and interpretation of the NGS outputs. In fact, this single challenge is mentioned in more than two-thirds of all the identified studies. The lack of large-scale validation of NGS outputs with regard to costs and data complexity is challenging and perhaps not feasible for individual research groups to achieve, hence its function as an

Table. A summary of the most commonly mentioned challenges, solutions, and implementation potentials for next-generation sequencing on the field of influenza virus research

Challenges	**References**
The need for complicated bioinformatics analysis as NGS delivers high volumes of raw reads	Deng *et al.* (2015), Cheval *et al.* (2011), Torres *et al.* (2008), Nowrousian (2010), Alquezar-Planas *et al.* (2013), Kampmann *et al.* (2011), Frey *et al.* (2014), Zhao *et al.* (2015), Lee *et al.* (2013), Archer *et al.* (2012), Bidzhieva *et al.* (2014), Kuroda *et al.* (2015), Iqbal *et al.* (2014), MacLean *et al.* (2009), Radford *et al.* (2012), Peng *et al.* (2014), Peng *et al.* (2011)
The high cost and less availability of NGS equipment	Fischer *et al.* (2015), Deng *et al.* (2015), Ansorge (2009), Zhao *et al.* (2015), MacLean *et al.* (2009)
Requirements for clinical assay validation	Fischer *et al.* (2015), Kampmann *et al.* (2011), Rutvisuttinunt *et al.* (2015), Frey *et al.* (2014)
Solutions	**References**
Clinical validation of NGS	Fischer *et al.* (2015)
Development of an automated assembly and analysis pipeline can make the bioinformatics analysis of transferring raw reads to the specific genomic identification more efficient	Alquezar-Planas *et al.* (2013), Frey *et al.* (2014)
Batching and multiplexing samples in single sequencing runs, while maintaining error rates and relative cost low	Ansorge (2009), Lee *et al.* (2013)
Implementation	**References**
Allows the full genome sequencing of influenza A viruses in a single run	Deng *et al.* (2015), Torres *et al.* (2008), Yu *et al.* (2014), Farsani *et al.* (2015), Lee *et al.* (2013), Téllez-Sosa *et al.* (2013), Archer *et al.* (2012), Zhou *et al.* (2014), Van Borm *et al.* (2012), Quail *et al.* (2012), Selleri (2013)
Generate an impressive amount of sequence information in a short time frame and high speed	Alquezar-Planas *et al.* (2013), Kampmann *et al.* (2011), Rutvisuttinunt *et al.* (2015), Farsani *et al.* (2015), Rutvisuttinunt *et al.* (2013), Flaherty *et al.* (2012), Téllez-Sosa *et al.* (2013), Archer *et al.* (2012), Bidzhieva *et al.* (2014), Leung *et al.* (2013), Watson *et al.* (2013), Kuroda *et al.* (2015), MacLean *et al.* (2009), Radford *et al.* (2012)
Has the potential to detect known and unknown pathogens (viruses, bacteria, fungi, and parasites), novel viruses in heterogeneous populations in a single application	Fischer *et al.* (2015), Nowrousian (2010), Lin *et al.* (2014), Alquezar-Planas *et al.* (2013), Ansorge (2009), Yu *et al.* (2014), Kampmann *et al.* (2011), Rutvisuttinunt *et al.* (2015), Frey *et al.* (2014), Van den Hoecke *et al.* (2015), Kuroda *et al.* (2015)

NGS = next-generation sequencing

adoptive impediment. The availability of NGS equipment is a second most popular challenge, followed by the high cost of the new technique compared with existing traditional methods.

The solutions suggested to overcome these issues were much more diverse and fragmented in nature. A large number of authors stressed the need for the development of an automated assembly and development software pipeline, making the whole NGS downstream analyses more efficient and reliable. Although most authors appreciate the production of a series of standard operating procedures, very few are willing to test (individually or institutionally) and compare the different recommended standard operating procedures. The ability to match and multiplex the samples in single sequencing runs is one of the solutions implemented to create cost efficiencies according to the manufacturers' recommendations.

The opportunities that NGS provides to research are evident to all authors. The ability to produce a number of complete influenza genomes in a single run at high resolution and the potential to detect heterogeneous populations in a single application are clearly outlined. The production of considerably larger amounts of sequence information in a short time frame and high speed as compared with traditional molecular methods was also welcome.

5. Discussion

In the last few years, high-throughput NGS technologies have become more widely available, and they are under continuous improvement and development. NGS has already been used in several projects, in metagenomics, whole genome sequencing, RNA sequencing, and small RNA discovery (Barzon *et al.* 2011). These technologies confer advantages over older methods, including single-molecule sequencing, high-throughput and increased quantity of sequencing data, while avoiding the necessity for cloning individual DNA fragments (Ansorge 2009).

However, NGS technologies share common features that still limit their use. Through the current search, these have been identified as being the generation of high-throughput data that require substantial computational resources for their subsequent analyses and quality control, the high comparative cost of sequencing using NGS, and the availability of suitable equipment (Deng *et al.* 2015; Metzker 2010). As such, the complete replacement of the Sanger-based methods is yet unlikely, until the aforementioned barriers are addressed successfully. The NGS cost per run and the cost per sample has already decreased substantially, and higher multiplexing approaches exert further pressure toward this direction (Quiñones-Mateu *et al.* 2014).

According to our current observations, the adoption of NGS sequencing in influenza research seems to correlate well with Buxton's law, where “it is always too early [for rigorous evaluation] until, unfortunately, it's suddenly too late (Buxton and Drummond 1987).” The initial adopters of NGS are unable or reluctant to apply formative assessment of the different existing technologies, in part because the technologies themselves are still under development. However, as the clinical introduction of NGS starts to materialize, the number of NGS adopters increases and the technique becomes more familiar and integrated within organized facilities, and the completion of an evidence-based assessment will be even more difficult to materialize.

In practice, the current NGS applications are very similar to most newly implemented innovations, composed of a hard core of fixed techniques (e.g. library preparation) with a soft periphery of features (e.g. bioinformatics analyses). The existence of this soft periphery means that the distribution of risk and benefits for the adopters is not entirely fixed as NGS can be implemented in a variety of ways that are not fully clarified by the existing peer-reviewed literature (Ilinca *et al.* 2012). The uncertainty surrounding some of the implementations and outputs would be expected to still generate a multitude of different claims and adoption pathways.

Having said that, NGS is a very successful platform for viral research studies as it has already led to the discovery of novel viruses and their association of pathogenesis in diseases (Quiñones-

Mateu *et al.* 2014). Hence, it is widely expected that these technologies will be applied in routine clinical virology laboratories for nearly all viral pathogens including influenza viruses in the not-so-distant future (Gibson *et al.* 2014; Swenson *et al.* 2011; Kagan *et al.* 2012).

Conflict of interest

The authors report no conflict of interest.

Acknowledgements

The authors acknowledge the contribution of Prof Andrew Hayward and Dr Laura Shallcross in the initial stages of the study preparation.

This publication presents independent research supported by the Health Innovation Challenge Fund T5-344 (ICONIC), a parallel funding partnership between the Department of Health and Wellcome Trust. The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health or Wellcome Trust.

References

Alquezar-Planas DE, Mourier T, Bruhn CA, Hansen AJ, Vitcetz SN, Mørk S, Gorodkin J, Nielsen HA, Guo Y, Sethuraman A, Paxinos EE, Shan T, Delwart EL, Nielsen LP. 2013. Discovery of a divergent HPIV4 from respiratory secretions using second and third generation metagenomic sequencing. *Sci Rep* 14:2468.

Ansorge WJ. 2009. Next-generation DNA sequencing techniques. *N Biotechnol* 25: 195–203.

Archer J, Baillie G, Watson SJ, Kellam P, Rambaut A, Robertson DL. 2012. Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. *BMC Bioinforma* 13:47.

Bainbridge MN, Warren RL, Hirst M, Romanuik T, Zeng T, Go A, Delaney A, Griffith M, Hickenbotham M, Magrini V, Mardis ER, Sadar MD, Siddiqui AS, Marra MA, Jones SJ. 2006. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics* 7:246.

Barzon L, Lavezzo E, Militello V, Toppo S, Palù G. 2011. Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci* 12: 7861–84.

Baum A, Sachidanandam R, Garcia-Sastre A. 2010. Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proc Natl Acad Sci U S A* 107:16303–8.

Bialasiewicz S, McVernon J, Nolan T, Lambert SB, Zhao G, Wang D, Nissen MD, Sloots TP. 2014. Detection of a divergent Parainfluenza 4 virus in an adult patient with influenza like illness using next-generation sequencing. *BMC Infect Dis* 14:275.

Bidzhieva B, Zagorodnyaya T, Karagiannis K, Simonyan V, Laassri M, Chumakov K. 2014. Deep sequencing approach for genetic stability evaluation of influenza A viruses. *J Virol Methods* 199:68–75.

Borozan I, Watt SN, Ferretti V. 2013. Evaluation of alignment algorithms for discovery and identification of pathogens using RNA-seq. *PLoS One* 8.

Buggele WA, Krause KE, Horvath CM. 2013. Small RNA profiling of influenza A virus-infected cells identifies miR-449b as a regulator of histone deacetylase 1 and interferon beta. *PLoS One* 8:76560.

Burnham AJ, Armstrong J, Lowen AC, Webster RG, Govorkova EA. 2015. Competitive fitness of influenza B viruses with neuraminidase inhibitor-resistant substitutions in a coinfection model of the human airway epithelium. *J Virol* 89: 4575–87.

Buxton MJ. 1987. Problems in the economic appraisal of new health technology: the evaluation of heart transplants in the UK. In: Drummond MF (Ed.). *Economic appraisal of health technology in the European Community*. Oxford: Oxford Medical Publications. pp. 103–18.

Cheval J, Sauvage V, Frangeul L, Dacheux L, Guigon G, Dumey N, Pariente K, Rousseaux C, Dorange F, Berthet N, Brisse S, Moszer I, Bourhy H, Manuguerra CJ, Lecuit M, Burguiere A, Caro V, Eloit M. 2011. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J Clin Microbiol* 49:3268–75.

Clavijo A, Nikooienejad A, Esfahani MS, Metz RP, Schwartz S, Atashpaz-Gargari E, Deliberto TJ, Lutman MW, Pedersen K, Bazan LR, Koster LG, Jenkins-Moore M, Swenson SL, Zhang M, Beckham T, Johnson CD, Bounpheng M. 2013. Identification and analysis of the first 2009 pandemic H1N1 influenza virus from U.S. feral swine. *Zoonoses Public Health* 60:327–35.

Deng YM, Spirason N, Iannello P, Jelley L, Lau H, Barr IG. 2015. A simplified sanger sequencing method for full genome sequencing of multiple subtypes of human influenza A viruses. *J Clin Virol* 68:43–8.

Depew J, Zhou B, McCorrison JM, Wentworth DE, Purushe J, Koroleva G, Fouts DE. 2013. Sequencing viral genomes from a single isolated plaque. *Virol J* 10:181.

Dugan VG, Dunham EJ, Jin G, Sheng ZM, Kaser E, Nolting JM, Alexander Jr HL, Slemons RD, Taubenberger JK. 2011. Phylogenetic analysis of low pathogenicity H5N1 and H7N3 influenza A virus isolates recovered from sentinel, free flying, wild mallards at one study site during 2006. *Virology* 417:98–105.

Dugan VG, Saira K, Ghedin E. 2012. Large-scale sequencing and the natural history of model human RNA viruses. *Future Virol* 7:563–73.

Farsani SM, Deijs M, Dijkman R, Molenkamp R, Jeeninga RE, Ieven M, Goossens H, van der Hoek L. 2015. Culturing of respiratory viruses in well-differentiated pseudostratified human airway epithelium as a tool to detect unknown viruses. *Influenza Other Respir Viruses* 9:51–7.

Fischer N, Indenbirken D, Meyer T, Lütgehetmann M, Lellek H, Spohn M, Aepfelbacher M, Alawi M, Grundhoff A. 2015. Evaluation of unbiased next-generation sequencing of RNA (RNA-seq) as a diagnostic method in influenza virus-positive respiratory samples. *J Clin Microbiol* 53:2238–50.

Flaherty P, Natsoulis G, Muralidharan O, Winters M, Buenrostro J, Bell J, Brown S, Holodniy M, Zhang N, Ji HP. 2012. Ultrasensitive detection of rare mutations using next-generation targeted resequencing. *Nucleic Acids Res* 40.

Frey KG, Herrera-Galeano JE, Redden CL, Luu TV, Servetas SL, Mateczun AJ, Mokashi VP, Bishop-Lilly KA. 2014. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics* 15:96.

Gibson RM, Meyer AM, Winner D, Archer J, Feyertag F, Ruiz-Mateos E, Leal M, Robertson DL, Schmotzer CL, Quiñones-Mateu ME. 2014. Sensitive deep sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism. *Antimicrob Agents Chemother* 58: 2167–85.

Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, Kim E, Pillai DR, Guyard C, Mazzulli T, Isa P, Arias CF, Hackett J, Schochetman G, Miller S, Tang P, Chiu CY. 2010. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One* 5:13381.

Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, et al.. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* 10.

Head MG, Fitchett JR, Nageshwaran V, Kumari N, Hayward AC, Atun R. 2016. Research investments in global health: a systematic analysis of UK infectious disease research funding and global health metrics, 1997–2013. *EBioMedicine* 3: 180–90.

Ilinca S, Hamer S, Botje D, Espin J, Veloso Mendes R, Mueller J, van Wijngaarden J, Vinot D, Plochg T. 2012. All you need to know about innovation in healthcare: the 10 best reads. Special issue: innovation in Healthcare. *Int J Healthc Manag* 5: 193–202.

International Human Genome Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–45.

Iqbal M, Reddy KB, Brookes SM, Essen SC, Brown IH, McCauley JW. 2014. Virus pathotype and deep sequencing of the HA gene of a low pathogenicity H7N1 avian influenza virus causing mortality in Turkeys. *PLoS One* 9.

Jakhesara SJ, Bhatt VD, Patel NV, Prajapati KS, Joshi CG. 2014. Isolation and characterization of H9N2 influenza virus isolates from poultry respiratory disease outbreak. *Springerplus* 3:196.

Jonges M, Welkers MR, Jeeninga RE, Meijer A, Schneeberger P, Fouchier RA, de Jong MD, Koopmans M. 2014. Emergence of the virulence-associated PB2 E627K substitution in a fatal human case of highly pathogenic avian influenza virus A(H7N7) infection as determined by Illumina ultra-deep sequencing. *J Virol* 88: 1694–702.

Kagan RM, Johnson EP, Siaw M, Biswas P, Chapman DS, Su Z, Platt JL, Pesano RL. 2012. A genotypic test for HIV-1 tropism combining Sanger sequencing with ultra deep sequencing predicts virologic response in treatment-experienced patients. *PLoS One* 7:46334.

Kampmann ML, Fordyce SL, Avila-Arcos MC, Rasmussen M, Willerslev E, Nielsen LP, Gilbert MT. 2011. A simple method for the parallel deep sequencing of full influenza A genomes. *J Virol Methods* 178:243–8.

Karlsson OE, Belák S, Granberg F. 2013. The effect of preprocessing by sequence-independent, single-primer amplification (SISPA) on metagenomic detection of viruses. *Biosecur Bioterror* 11:S227–34.

Kuroda M, Katano H, Nakajima N, Tobiume M, Ainai A, Sekizuka T, Hasegawa H, Tashiro M, Sasaki Y, Arakawa Y, Hata S, Watanabe M, Sata T. 2010. Characterization of quasi species of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PLoS One* 5: 10256.

Kuroda M, Niwa S, Sekizuka T, Tsukagoshi H, Yokoyama M, Ryo A, Sato H, Kiyota N, Noda M, Kozawa K, Shirabe K, Kusaka T, Shimojo N, Hasegawa S, Sugai K, Obuchi M, Tashiro M, Oishi K, Ishii H, Kimura H. 2015. Molecular evolution of the VP1, VP2, and VP3 genes in human rhinovirus species C. *Sci Rep* 5:8185.

Lange J, Groth M, Schlegel M, Krumbholz A, Wieczorek K, Ulrich R, Köppen S, Schulz K, Appl D, Selbitz HJ, Sauerbrei A, Platzer M, Zell R, Dürrwald R. 2013. Reassortants of the pandemic (H1N1) 2009 virus and establishment of a novel porcine H1N2 influenza virus, lineage in Germany. *Vet Microbiol* 167: 345–56.

Lee HK, Tang JW, Kong DH, Koay ES. 2013. Simplified large-scale Sanger genome sequencing for influenza A/H3N2 virus. *PLoS One* 8:64785.

León AJ, Banner D, Xu L, Ran L, Peng Z, Yi K, Chen C, Xu F, Huang J, Zhao Z, Lin Z, Huang SH, Fang Y, Kelvin AA, Ross TM, Farooqui A, Kelvin DJ. 2013. Sequencing,

annotation, and characterization of the influenza ferret infectome. *J Virol* 87: 1957–66.
Leung RK, Dong ZQ, Sa F, Chong CM, Lei SW, Tsui SK, Lee SM. 2013. Quick, sensitive and specific detection and evaluation of quantification of minor variants by high-throughput sequencing. *Mol Biosyst* 10:206–14.
Lin Z, Farooqui A, Li G, Wong GK, Mason AL, Banner D, Kelvin AA, Kelvin DJ, León AJ. 2014. Next-generation sequencing and bioinformatic approaches to detect and analyze influenza virus in ferrets. *J Infect Dev Ctries* 8:498–509.
Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ. 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* 10: 599–606.
MacLean D, Jones JD, Studholme DJ. 2009. Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* 7:287–96.
Metzker ML. 2010. Sequencing technologies – the next generation. *Nat Rev Genet* 11:31–46.
Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T. 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 4:4219.
Nowrousian M. 2010. Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryot Cell* 9:1300–10.
Park JY, Kricka LJ, Fortina P. 2013. Next-generation sequencing in the clinic. *Nat Biotechnol* 31:990–2.
Peng X, Alföldi J, Gori K, Eisfeld AJ, Tyler SR, Tisoncik-Go J, Brawand D, Law GL, Skunca N, Hatta M, Gasper DJ, Kelly SM, Chang J, Thomas MJ, Johnson J, Berlin AM, Lara M, Russell P, Swofford R, Turner-Maier J, Young S, Hourlier T, Aken B, Searle S, Sun X, Yi Y, Suresh M, Tumpey TM, Siepel A, Wisely SM, Dessimoz C, Kawaoka Y, Birren BW, Lindblad-Toh K, Di Palma F, Engelhardt JF, Palermo RE, Katze MG. 2014. The draft genome sequence of the ferret (*Mustela putorius furo*) facilitates study of human respiratory disease. *Nat Biotechnol* 32: 1250–5.
Peng X, Gralinski L, Ferris MT, Frieman MB, Thomas MJ, Proll S, Korth MJ, Tisoncik JR, Heise M, Luo S, Schroth GP, Tumpey TM, Li C, Kawaoka Y, Baric RS, Katze MG. 2011. Integrative deep sequencing of the mouse lung transcriptome reveals differential expression of diverse classes of small RNAs in response to respiratory virus infection. *mBio* 2.
Pettersson E, Zajac P, Ståhl PL, Jacobsson JA, Fredriksson R, Marcus C, Schiöth HB, Lundeberg J, Ahmadian A. 2008. Allelotyping by massively parallel pyrosequencing of SNP-carrying trinucleotide threads. *Hum Mutat* 29:323–9.
Power PM, Bentley SD, Parkhill J, Moxon ER, Hood DW. 2012. Investigations into genome diversity of Haemophilus influenzae using whole genome sequencing of clinical isolates and laboratory transformants. *BMC Microbiol* 12:273.
Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent. Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341.
Quiñones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA. 2014. Deep sequencing: becoming a critical tool in clinical virology. *J Clin Virol* 61:9–19.
Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. 2012. Application of next-generation sequencing technologies in virology. *J Gen Virol* 93:1853–68.
Ren X, Yang F, Hu Y, Zhang T, Liu L, Dong J, Sun L, Zhu Y, Xiao Y, Li L, Yang J, Wang J, Jin Q. 2013. Full genome of influenza A (H7N9) virus derived by direct sequencing without culture. *Emerg Infect Dis* 19.
Rutvisuttinunt W, Chinnawirotpisan P, Simasathien S, Shrestha SK, Yoon IK, Klungthong C, Fernandez S. 2013. Simultaneous and complete genome sequencing of influenza A and B with high coverage by Illumina MiSeq platform. *J Virol Methods* 193:394–404.
Rutvisuttinunt W, Chinnawirotpisan P, Thaisomboonsuk B, Rodpradit P, Ajariyakhajorn C, Manasatienkij W, Simasathien S, Shrestha SK, Yoon IK, Klungthong C, Fernandez S. 2015 Jul. Viral subpopulation diversity in influenza virus isolates compared to clinical specimens. *J Clin Virol* 68:16–23.
Saira K, Lin X, DePasse JV, Halpin R, Twaddle A, Stockwell T, Angus B, Cozzi-Lepri A, Delfino M, Dugan V, Dwyer DE, Freiberg M, Horban A, Losso M, Lynfield R, Wentworth DN, Holmes EC, Davey R, Wentworth DE, Ghedin E, INSIGHT FLU002 Study Group, INSIGHT FLU003 Study Group. 2013. Sequence analysis of in vivo defective interfering-like RNA of influenza A H1N1 pandemic virus. *J Virol* 87:8064–74.
Satkoski JA, Malhi R, Kanthaswamy S, Tito R, Malladi V, Smith D. 2008. Pyrosequencing as a method for SNP identification in the rhesus macaque (*Macaca mulatta*). *BMC Genomics* 9:256.
Selleri M. 2013. Detection of haemagglutinin D222 polymorphisms in influenza A(H1N1) pdm09-infected patients by ultra-deep pyrosequencing. *Clin Microbiol Infect* 19:668–73.
Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135–45.
Sikora D, Rocheleau L, Brown EG, Pelchat M. 2014. Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. *Sci Rep* 4:6181.
Swaminathan S, Hu X, Zheng X, Kriga Y, Shetty J, Zhao Y, Stephens R, Tran B, Baseler MW, Yang J, Lempicki RA, Huang D, Lane HC, Imamichi T. 2013. Interleukin-27 treated human macrophages induce the expression of novel microRNAs which may mediate anti-viral properties. *Biochem Biophys Res Commun* 434:228–34.
Swenson LC, Mo T, Dong WW, Zhong X, Woods CK, Jensen MA, Thielen A, Chapman D, Lewis M, James I, Heera J, Valdez H, Harrigan PR. 2011. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. *J Infect Dis* 203:237–45.
Tan YC, Blum LK, Kongpachith S, Ju CH, Cal X, Lindstrom TM, Sokolove J, Robinson WH. 2014. High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. *Clin Immunol* 151:55–65.
Téllez-Sosa J, Rodríguez MH, Gómez-Barreto RE, Valdovinos-Torres H, Hidalgo AC, Cruz-Hervert P, Luna RS, Carrillo-Valenzo E, Ramos C, García-García L, Martínez-Barnetche J. 2013. Using high-throughput sequencing to leverage surveillance of genetic diversity and oseltamivir resistance: a pilot study during the 2009 influenza A(H1N1) pandemic. *PLoS One* 8.
Torres TT, Metta M, Ottenwälder B, Schlötterer C. 2008. Gene expression profiling by massively parallel sequencing. *Genome Res* 18:172–7.
Tsai KN, Chen GW. 2011. Influenza genome diversity and evolution. *Microbes Infect* 13:479–88.
Van Borm S, Rosseel T, Vangeluwe D, Vandenbussche F, van den Berg T, Lambrecht B. 2012. Phylogeographic analysis of avian influenza viruses isolated from Charadriiformes in Belgium confirms intercontinental reassortment in gulls. *Arch Virol* 157:1509–22.
Van den Hoecke S, Verhelst J, Vuylsteke M, Saelens X. 2015. Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. *BMC Genomics* 16:79.
Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, García-Sastre A, tenOever BR. 2014. Influenza a virus transmission bottlenecks are defined by infection route and recipient host. *Cell Host Microbe* 16:691–700.
Wang Y, Brahmakshatriya V, Lupiani B, Reddy SM, Soibam B, Benham AL, Gunaratne P, Liu HC, Trakooljul N, Ing N, Okimoto R, Zhou H. 2012. Integrated analysis of microRNA expression and mRNA transcriptome in lungs of avian influenza virus infected broilers. *BMC Genomics* 13:278.
Watson SJ, Welkers MR, Depledge DP, Coulter E, Breuer JM, de Jong MD, Kellam P. 2013. Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos Trans R Soc Lond B Biol Sci* 368.
Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452:872–6.
Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, Baker D. 2012. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* 30:543–8.
Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Chen SH, Lu IH, Lin CY, Chin RG, Luan HH, Nguyen N, Nelson SF, Li X, Wu TT, Sun R. 2014. High-throughput profiling of influenza a virus hemagglutinin gene at single-nucleotide resolution. *Sci Rep* 4.
Xiao YL, Kash JC, Beres SB, Sheng ZM, Musser JM, Taubenberger JK. 2013. High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. *J Pathol* 229:535–45.
Yasugi M, Nakamura S, Daidoji T, Kawashita N, Ramadhany R, Yang CS, Yasunaga T, Iida T, Horii T, Ikuta K, Takahashi K, Nakaya T. 2012. Frequency of D222G and Q223R hemagglutinin mutants of pandemic (H1N1) 2009 influenza virus in Japan between 2009 and 2010. *PLoS One* 7.
Yu X, Jin T, Cui Y, Pu X, Li J, Xu J, Liu G, Jia H, Liu D, Song S, Yu Y, Xie L, Huang R, Ding H, Kou Y, Zhou Y, Wang Y, Xu X, Yin Y, Wang J, Guo C, Yang X, Hu L, Wu X, Wang H, Liu J, Zhao G, Zhou J, Pan J, Gao GF, Yang R, Wang J. 2014. Influenza H7N9 and H9N2 viruses: coexistence in poultry linked to human H7N9 infection and genome characteristics. *J Virol* 88:3423–31.
Zhao J, Ragupathy V, Liu J, Wang X, Vemula SV, El Mubarak HS, Ye Z, Landry ML, Hewlett I. 2015. Nanomicroarray and multiplex next-generation sequencing for simultaneous identification and characterization of influenza viruses. *Emerg Infect Dis* 21:400–8.
Zhirnov OP, Vorobjeva IV, Saphonova OA, Poyarkov SV, Ovcharenko AV, Anhlan D, Malyshev NA. 2009. Structural and evolutionary characteristics of HA, NA, NS and M genes of clinical influenza A/H3N2 viruses passaged in human and canine cells. *J Clin Virol* 45:322–33.
Zhou B, Lin X, Wang W, Halpin RA, Bera J, Stockwell TB, Barr IG, Wentworth DE. 2014. Universal influenza B virus genomic amplification facilitates sequencing, diagnostics, and reverse genetics. *J Clin Microbiol* 52:1330–7.