The Journal of Infectious Diseases

SUPPLEMENT ARTICLE

Next-Generation Human Immunodeficiency Virus Sequencing for Patient Management and Drug Resistance Surveillance

Marc Noguera-Julian,[1,2,3] Dianna Edgil,[4] P. Richard Harrigan,[5] Paul Sandstrom,[6] Catherine Godfrey,[7] and Roger Paredes[1,2,3,8]

[1]IrsiCaixa AIDS Research Institute, Badalona, [2]Universitat Autònoma de Barcelona, Bellaterra, and [3]Universitat de Vic-Universitat Central de Catalunya, Catalonia, Spain; [4]United States Agency for International Development, Washington, District of Columbia; [5]British Columbia Centre for Excellence in HIV/AIDS, University of British Columbia, Vancouver, and [6]National HIV and Retrovirology Laboratory, JC Wilt Infectious Diseases Research Centre, Public Health Agency of Canada, Winnepeg, Manitoba, Canada; [7]Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland; and [8]Infectious Diseases Unit, Hospital Universitari Germans Trias i Pujol, Badalona, Catalonia, Spain

High-quality, simplified, and low-cost human immunodeficiency virus (HIV) drug resistance tests that are able to provide timely actionable HIV resistance data at individual, population, and programmatic levels are needed to confront the emerging drug-resistant HIV epidemic. Next-generation sequencing technologies embedded in automated cloud-computing analysis environments are ideally suited for such endeavor. Whereas NGS can reduce costs over Sanger sequencing, automated analysis pipelines make NGS accessible to molecular laboratories regardless of the available bioinformatic skills. They can also produce highly structured, high-quality data that could be examined by healthcare officials and program managers on a real-time basis to allow timely public health action. Here we discuss the opportunities and challenges of such an approach.

Keywords. HIV resistance; NGS; bioinformatics; surveillance; public health.

Public health–oriented approaches to global antiretroviral therapy (ART) delivery have achieved striking reductions in human immunodeficiency virus (HIV)/AIDS mortality, morbidity, and transmissions during the last decade [1, 2]. However, the emergence of HIV drug resistance to currently recommended drugs—mostly nonnucleoside reverse transcriptase inhibitors (NNRTIs), but also to nucleos(t)ide reverse transcriptase inhibitors—in several low- and middle-income countries (LMICs) is posing a threat to the future control of the HIV pandemic [3–5].

In response to this growing problem, the World Health Organization (WHO) has released guidance suggesting changing to dolutegravir-based first-line ART in LMICs with pretreatment NNRTI resistance rates >10% [6]. In countries where such transition may not be readily feasible, patient-level drug resistance testing might be considered an option to guide first-line ART. It is thus likely that HIV drug resistance testing will be gradually adopted, albeit with different speed and at different levels, by various LMICs during the next decade. This poses a major challenge to already overburdened healthcare systems currently struggling to ensure widespread use of viral load testing for HIV clinical management.

HIV genotyping in resource-rich countries is aimed at personalizing ART and maximizing its efficacy for each patient. In contrast, ART in LMICs must be provided under a public health approach that aims to use one simple regimen for all individuals on therapy. Epidemiological information about regional ART resistance patterns is used to design a single national regimen. This allows for efficiencies of scale in manufacture, purchasing, and distribution. HIV drug resistance testing in LMICs must not only provide patient-level information, but also must provide robust, epidemiological information that can inform meaningful public action in a timely fashion. Drug resistance testing should provide high-quality, well-structured data to public health officials to detect emerging trends in resistance and transmission in populations of interest. Next-generation sequencing (NGS) technologies are ideally suited for such endeavor, because they can be embedded in a cloud-computing environment that enables automated, robust, high-quality data analysis at individual as well as population and programmatic levels.

NGS platforms are evolving rapidly (reviewed in [7]), and all produce large amounts of sequence data. With sufficient sample multiplexing, NGS can provide HIV genotypes for <$50 per sample, improving its cost effectiveness [8, 9]. However, automatized bioinformatics pipelines are needed because tens to thousands of sequences per sample must be analyzed [8, 10–12]. Many laboratories in resource-rich settings are transitioning from Sanger sequencing to NGS. NGS is equally robust and laboratory demanding, but provides additional data on low-frequency HIV drug-resistant mutants [13] (discussed by

Correspondence: R. Paredes, MD, PhD, Infectious Diseases Unit and irsiCaixa AIDS Research Institute, Hospital Universitari Germans Trias i Pujol, Ctra de Canyet s/n, 08916 Badalona, Catalonia, Spain (rparedes@irsicaixa.es).

The Journal of Infectious Diseases® 2017;XX00:S1–5

DOI: 10.1093/infdis/jix397

Stella-Ascariz et al in this issue) and, most importantly, allows sharp per-test cost reductions. In addition, NGS is suitable for different HIV subtypes [14], can be performed in dried blood spots [15–17], and can provide information on longer HIV genome fragments, including near–full genome sequences, which improves the resolution of phylogenetic and phylogeographic analyses to track the HIV epidemic.

The need for bioinformatic analyses has traditionally been a major drawback for NGS implementation in routine clinical practice. However, this can also become its major asset. Most bioinformatic software tools have been developed within the HIV research context and therefore demand local computational infrastructure or bioinformatics skills [18, 19]. Bioinformatic knowledge and computational infrastructure are usually not available in most LMICs' diagnostic laboratories, and the setup of these capabilities in such environments is an important handicap for NGS implementation [20]. In recent years, there have been major advances to make NGS analysis pipelines highly automated, streamlined, and simpler to use, to the extent that some of them can be used by laboratory technicians without bioinformatic skills (Table 1). Such automated pipelines usually include filtering of low-quality data, alignment of short sequences against an HIV reference sequence, and quantification of amino acid variants present in the viral sample, with resistance mutations above a prespecified abundance threshold used for HIV resistance interpretation [21].

Figure 1 summarizes a putative simplified NGS analysis workflow foreseen for LMICs. Upon a clinician's or program officer's request, a central laboratory would run an NGS HIV resistance test. The laboratory technician would upload the obtained raw NGS data files into a cloud-based internet resource and select a prespecified analysis configuration (eg, HIV gene, sequencing depth, HIV resistance interpretation algorithm, sequence quality settings). The analyses would be run automatically in the cloud in a hands-off process that also enables local computers to be turned off or used for other purposes. After 10–30 minutes, the analyses would be ready, and a patient-level HIV resistance report automatically interpreted with the Stanford HIVdb (or equivalent) rules would be retrieved, alongside an automated prespecified assessment of genotyping test quality (eg, sequence coverage, sequence quality, contamination, cross-sample contamination). This report would be sent back to the clinician/program officer to act upon it. Data would remain stored in a highly structured manner for further population-level analyses, which may include quality monitoring at a laboratory, center, and/or program levels, and real-time surveillance of transmitted and emerging HIV resistance including the identification of HIV resistance hotspots and transmission clusters of phylogenetically related viruses. Standard security-compliant encryption and authentication mechanisms on stored data would allow definition of region/institution/user-specific data access policies and, when needed, sharing among different institutions.

Ideally, automated NGS analysis software should (1) be remotely usable by users with no bioinformatics skills through a user-friendly web interface accessible from simple computers or smartphones; (2) provide robust, reproducible, and easy-to-interpret results using standard and well-established HIV resistance interpretation rules (eg, HIVdb or equivalent); (3) incorporate built-in quality standards; (4) avoid unnecessary transfer of large data volumes and provide clinically actionable results that can be downloadable with limited network access; (5) demand minimal or no local computational infrastructure; (6) seamlessly respond to varying number of samples in a highly scalable manner without an impact in time to results; and (7) have minimal cost to enable their sustainable adoption by LMICs. Some available software (Table 1)

Table 1. Automated Next-Generation Sequencing Analysis Pipelines for Human Immunodeficiency Virus Resistance

Software	URL	Cost / Sample	Time	Bioinformatic–IT Needs[b]	Cloud Based	Year	Commercial Use	Web Interface
V-Phaser [19]	https://www.broadinstitute.org/viral-genomics/v-phaser-2	Free		Yes	No	2013	No	No
ShoRah [26]	https://github.com/cbg-ethz/shorah	Free		Yes	No	2013	No	No
VirVarSeq [27]	https://sourceforge.net/projects/virtools/	Free		Yes	No	2015	No	No
VirFlow [28]	https://github.com/ICONIC-UCL/vir-flow	Free	3 h	Yes	Yes[c]	2016	No	No
MinVar[a] [29]	https://ozagordi.github.io/MinVar/	Free	<1 h	Yes	Yes[c]	2016	No	No
MiCall[a]	https://github.com/cfe-lab/MiCall	Free		No	Yes	2016	No	Yes
HyDRA	https://hydra.canada.ca/	Free	<1 h	No	Yes	2016	No	Yes
PASeq[a]	https://www.paseq.org	Free	<1 h	No	Yes	2016	No	Yes
SmartGeneHIV	http://www.smartgene.com/mod_ngs.html	NA		No	No	2016	Yes	No
DeepChekHIV [30]	https://www.ablsa.com/overview/deepchek/	$65	<1 h	No	Yes	2014	Yes	Yes

Many software tools exists that can serve for analysis of HIV drug resistance from NGS data. Their name and URL (last accessed April 2017) are provided. DeepChekHIV, PASeq, MiCall and Hydra are web-accessible cloud-based systems. MinVar and VirFlow can be ported to a cloud-based system using standard computational tools. PASeq and MiCall will be accessible from Illumina's Basespace.

Abbreviations: IT, information technology; NA, not available.

[a]Actively being developed/used for human immunodeficiency virus next-generation sequencing genotyping.

[b]Refers to the need of on-site computational infrastructure or expert staff.

[c]Can be ported to Cloud.

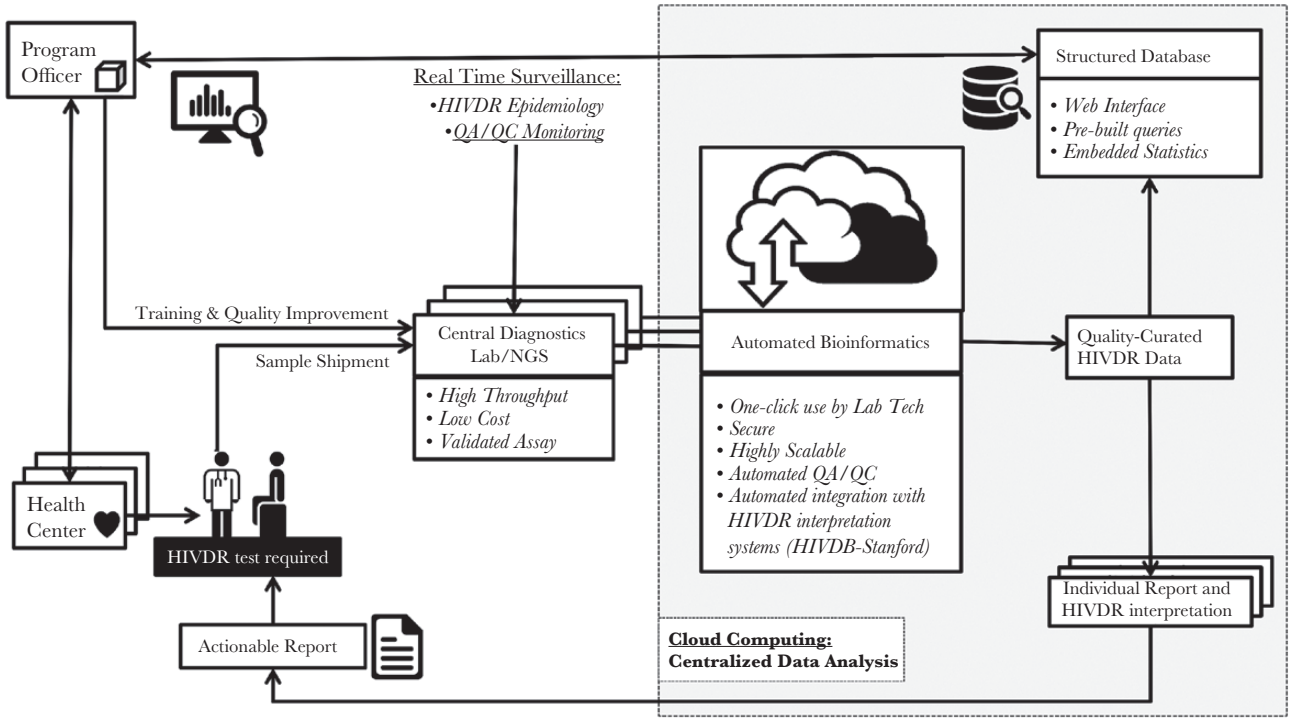

Figure 1. Next-generation sequencing analysis workflow in low- and middle-income countries. Abbreviations: HIVDR, human immunodeficiency virus drug resistance; NGS, next-generation sequencing; QA/QC, quality assurance/quality control.

can fulfill some or many of these conditions with little further development. Cloud-based software will facilitate scalability, data structure, and storage [22, 23]. Analysis parameters will be predefined by health programs expanding NGS testing to general population at a local level and seamlessly introduced, along with quality assurance (QA)/quality control (QC) thresholds into one-click actionable pipelines, but local-level policies for NGS data management and access still need to be defined. As a real-life example, analysis of 11 HIV-1 genotype samples using 3 already existing free software pipelines (PaSeq, MiCall, and Hydra) showed high level of agreement in resistance detection between them, even when low-frequency variants were evaluated Figure 2.

Successful implementation of HIV drug resistance testing with either Sanger or NGS in the real world, however, will be dependent upon several factors. The introduction of new technologies into low-resource settings presents challenges at the policy, laboratory, clinical, and systems levels [24]. Lessons learned from the scale-up of molecular testing for HIV and TB indicate a slow path to success in the absence of international funding and strong donor endorsement of programmatic implementation of new technologies [24, 25]. Rapidly growing access to viral load monitoring and the existence of practicable technologies, such as NGS, to detect drug-resistant HIV make HIV drug resistance testing for individual patient management a critical factor in supporting the rational use of available drug regimens in the longer-term.

The incorporation of routine HIV drug resistance testing into WHO and national HIV treatment guidelines, and further adoption into country testing algorithms, will precede implementation and is under discussion at the present time for specific high-risk populations. Prior to implementation, diagnostic algorithms, laboratory information systems, patient tracking and notification systems, and methods for monitoring and assessment of HIV programs will need to be revised using clinical indicators for HIV drug resistance. Increased capacity for identifying patients requiring second- and third-line antiretrovirals must be matched by the presence of appropriate antiretroviral regimens in countries performing HIV drug resistance testing. Conversely, the implementation of diagnostics for surveillance purposes has fewer associated challenges than for patient management, setting the stage for phased scale-up of testing over time. In fact, some level of HIV drug resistance surveillance is in place in many LMICs in sub-Saharan Africa.

At a systems level, issues regarding manufacturing capacity of the producers of new technologies, as well as rapid increases in the manufacture of familiar tests, can result in product shortages, requiring slower than anticipated scale-up, or decreases in quality, resulting in recalls. Repair capacity for equipment and parts supply needs to be available and in-country before expensive equipment purchase. Supply chain, sample transport, and results return are some of the most critical challenges facing molecular diagnostics today. Depending upon the handling specifications of the finished technology, cold chain requirements for reagents,

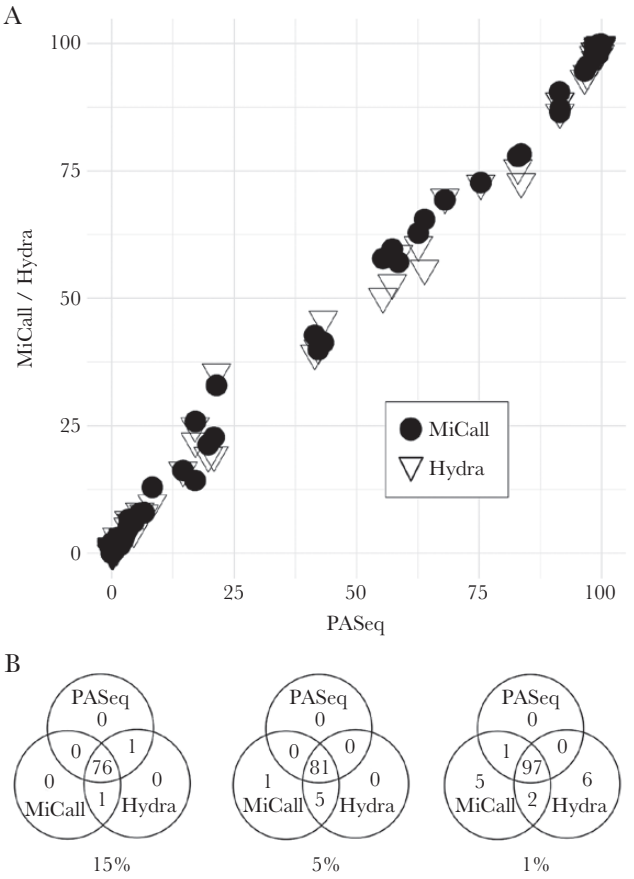

Figure 2. Agreement between 3 publicly available next-generation sequencing analysis pipelines in the analysis of 11 subtype C human immunodeficiency virus type 1 patient samples. *A*, Linearity in mutant frequency measurements between 1% and 100% mutant frequency. *B*, Venn diagram of mutants detected by each pipeline at 15%, 5%, and 1% thresholds.

bulky consumables, and short shelf lives complicate the supply chain's ability to ensure an uninterrupted supply of commodities for HIV drug resistance testing and require updating or expansion of storage facilities at the warehouse and laboratory.

Understanding the placement of NGS within the context of the overall HIV diagnostics network and the absorptive capacity of individual laboratories is essential. As treatment programs implement new paradigms of therapy, there is a tension between devolving laboratory testing to a local level and centralizing the more specialized testing. (Point-of-care resistance testing is discussed elsewhere in this supplement.) NGS, as a medium- to high-throughput diagnostic test with significant data capture and analytics requirements similar to most HIV drug resistance testing technologies, is currently best suited for a centralized testing network using higher-tier laboratories [8]. This approach supports both ongoing surveillance needs as well as the scope of diagnostics needed within the patient population (ie, diagnostic following confirmed treatment failure). Centralizing or partially centralizing testing to higher-functioning facilities can alleviate some of the systems-level issues, but may place additional burden on sample transport and results return networks. The latter of these, for many LMICs without strong return networks, largely depend upon the return of paper-based reports from the testing laboratory to the referring clinician and then back to the patient. This approach will also avert infrastructure challenges (eg, intermittent electrical supply, high temperatures, lack of network connection) that can result in service disruption.

Laboratory training, even for higher-tier laboratories familiar with HIV drug resistance testing for surveillance, will require training on the implementation of HIV drug resistance testing for clinical diagnosis. Expansion from surveillance to individual patient management will require clinician and laboratory training, and patient sensitization to ensure supply and demand of HIV drug resistance testing.

Finally, the need for external quality assurance of clinical laboratories and procedures and regular calibration of instruments, including regular access to service maintenance, is critical for the continued operation of HIV drug resistance testing instruments. The adaption of QC protocols routinely used by clinical chemistry laboratories, the consistent use of controls, and monitoring and review of data to identify shifts or trends can assist the early identification of issues with testing quality. Automated analysis pipelines could drastically reduce QA/QC problems derived from NGS data analysis and interpretation.

Next-generation sequencing thus provides a unique opportunity to achieve high-quality, low-cost HIV genotyping data useful for individual, population, and program-based HIV drug resistance assessments and real-time HIV resistance surveillance. The commitment of global health agencies, the molecular diagnostics industry, and LMICs to overcome the implementation challenges ahead will prove essential to achieve a durable control—and hopefully end—the HIV/AIDS pandemic.

Notes

Disclaimer. The views expressed in this article do not necessarily represent those of the National Institutes of Health, the US Agency for International Development, or the US government.

Financial support. R. P. and M. N.-J. are partially supported by Gala SIDA 2015 and 2016 editions, People in Red Gala 2016 edition, and the Fundació Glòria Soler.

Supplement sponsorship. This supplement was sponsored by the National Institute of Allergy and Infectious Disease, NIH, and the Centers for Disease Control and Prevention.

Potential conflicts of interest. R. P. has consulted for MSD and ViiV, and received research support from MSD and ViiV. All other authors report no potential conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. World Health Organization. HIV drug resistance. Global report on early warning indicators of HIV drug resistance. **2016**.
2. Clutter DS, Jordan MR, Bertagnolio S, Shafer RW. HIV-1 drug resistance and resistance testing. Infect Genet Evol **2016**; 46:292–307.
3. Bissio E, Barbás MG, Bouzas MB, et al. Pretreatment HIV-1 drug resistance in Argentina: results from a surveillance study performed according to WHO-proposed new methodology in 2014–15. J Antimicrob Chemother **2017**; 72:504–10.

4. Ávila-Ríos S, García-Morales C, Matías-Florentino M, et al. Pretreatment HIV-drug resistance in Mexico and its impact on the effectiveness of first-line antiretroviral therapy: a nationally representative 2015 WHO survey. Lancet HIV **2016**; 3:e579–91.
5. Casadellà M, Noguera-Julian M, Sunpath H, et al. Treatment options after virological failure of first-line tenofovir-based regimens in South Africa: an analysis by deep sequencing. AIDS **2016**; 30:1137–40.
6. World Health Organization. Guidelines on the public health response to pretreatment HIV drug resistance. Geneva, Switzerland: WHO, **2017**.
7. Casadellà M, Paredes R. Deep sequencing for HIV-1 clinical management. Virus Res **2016**. doi:10.1016/j.virusres.2016.10.019.
8. Inzaule SC, Ondoa P, Peter T, et al. Affordable HIV drug-resistance testing for monitoring of antiretroviral therapy in sub-Saharan Africa. Lancet Infect Dis **2016**; 16:e267–75.
9. Levison JH, Wood R, Scott CA, et al. The clinical and economic impact of genotype testing at first-line antiretroviral therapy failure for HIV-infected patients in South Africa. Clin Infect Dis **2013**; 56:587–97.
10. Lapointe H, Dong W, Lee GQ, et al. HIV drug resistance testing by high-multiplex "wide" sequencing on the Illumina MiSeq. Antimicrob Agents Chemother **2015**; 59:6824–33.
11. Dudley DM, Chin EN, Bimber BN, et al. Low-cost ultra-wide genotyping using Roche/454 pyrosequencing for surveillance of HIV drug resistance. PLoS One **2012**; 7:e36494.
12. Ekici H, Rao SD, Sönnerborg A, Ramprasad VL, Gupta R, Neogi U. Cost-efficient HIV-1 drug resistance surveillance using multiplexed high-throughput amplicon sequencing: implications for use in low- and middle-income countries. J Antimicrob Chemother **2014**; 69:3349–55.
13. Cozzi-Lepri A, Noguera-Julian M, Di Giallonardo F, et al; CHAIN Minority HIV-1 Variants Working Group. Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: a multicohort European case-control study using centralized ultrasensitive 454 pyrosequencing. J Antimicrob Chemother **2015**; 70:930–40.
14. Dudley DM, Bailey AL, Mehta SH, et al. Cross-clade simultaneous HIV drug resistance genotyping for reverse transcriptase, protease, and integrase inhibitor mutations by Illumina MiSeq. Retrovirology **2014**; 11:122.
15. Ji H, Li Y, Graham M, et al. Next-generation sequencing of dried blood spot specimens: a novel approach to HIV drug-resistance surveillance. Antivir Ther **2011**; 16:871–8.
16. Hollegaard MV, Grove J, Thorsen P, Nørgaard-Pedersen B, Hougaard DM. High-throughput genotyping on archived dried blood spot samples. Genet Test Mol Biomarkers **2009**; 13:173–9.
17. Salazar-Gonzalez JF, Salazar MG, Tully DC, et al. Use of dried blood spots to elucidate full-length transmitted/founder HIV-1 genomes. Pathog Immun **2016**; 1:129–53.
18. Kijak GH, Pham P, Sanders-Buell E, et al. Nautilus: a bioinformatics package for the analysis of HIV type 1 targeted deep sequencing data. AIDS Res Hum Retroviruses **2013**; 29:1361–4.
19. Yang X, Charlebois P, Macalalad A, Henn MR, Zody MC. V-Phaser 2: variant inference for viral populations. BMC Genomics **2013**; 14:674.
20. Beerenwinkel N, Günthard HF, Roth V, Metzner KJ. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front Microbiol **2012**; 3:329.
21. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think! Genome Biol **2011**; 12:125.
22. Haskew J, Rø G, Saito K, et al. Implementation of a cloud-based electronic medical record for maternal and child health in rural Kenya. Int J Med Inform **2015**; 84:349–54.
23. Mulder NJ, Adebiyi E, Adebiyi M, et al. Development of bioinformatics infrastructure for genomics research in H3Africa [manuscript published online ahead of print 13 March 2017]. Glob Heart **2017**. doi:10.1016/j.gheart.2017.01.005.
24. Médecins Sans Frontières. Making viral load routine. Successes and challenges in the implementation of routine HIV viral load monitoring. Part 1: programmatic strategies. Geneva, Switzerland: MSF, **2016**.
25. Lawn SD, Mwaba P, Bates M, et al. Advances in tuberculosis diagnostics: the Xpert MTB/RIF assay and future prospects for a point-of-care test. Lancet Infect Dis **2013**; 13:349–61.
26. Osvaldo Zagordi AB. ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data. BMC Bioinformatics **2011**; 12:119.
27. Verbist BMP, Thys K, Reumers J, et al. VirVarSeq: a low-frequency virus variant detection pipeline for Illumina sequencing using adaptive base-calling accuracy filtering. Bioinformatics **2015**; 31:94–101.
28. Cassarino TG, Frampton D, Sugar R, Charles E, Zisis Kozlakidis PK. High-throughput pipeline for de-novo assembly and drug resistance mutations identification from next-generation sequencing viral data of residual diagnostic samples. bioRxiv **2016**; doi: 10.1101/035154.
29. Huber M, Metzner KJ, Geissberger FD, et al. MinVar: a rapid and versatile tool for HIV-1 drug resistance genotyping by deep sequencing. J Virol Methods **2017**; 240:7–13.
30. Garcia-Diaz A, McCormick A, Booth C, et al. Analysis of transmitted HIV-1 drug resistance using 454 ultra-deep-sequencing and the DeepChek(*)-HIV system. J Int AIDS Soc **2014**; 17:19752.