Beyond surveillance: leveraging the potential of next generation sequencing in clinical virology

Paula C. Aulicino[1,2]* and Jason T. Kimata[3]

[1]Área de Laboratorios Especializados, Hospital de Pediatría "Prof. Dr. Juan P. Garrahan", Buenos Aires, Argentina, [2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina, [3]Department of Molecular Virology & Microbiology, Baylor College of Medicine, Houston, TX, United States

The COVID-19 pandemic demonstrated the strength of massive sequencing or next generation sequencing (NGS) techniques in viral genomic characterization. Millions of complete SARS-CoV-2 genomes were sequenced in almost real time. Laboratories around the world dedicated to the molecular diagnosis of infectious diseases were equipped with cutting-edge technologies for deep sequencing, coupled with strengthening or development of previously limited bioinformatics capacities. Almost 5 years have passed from the initiation of the COVID pandemic, and, opposed to what could be envisioned as an opportunity for viral genomics to expand, this has essentially scaled back in most clinical settings. Most of NGS equipment and capacities in many regions of the world have been repurposed for the study of cancer driver mutations, microbiome-related diseases, and pharmacogenomics, as the most important applications in health. Although financial constraints can limit their implementation, technical, regulatory, medical and data management factors are also part of the equation that will or will not make NGS a real game changer for advancing healthcare and guiding clinical decisions related to viral infections.

Next generation sequencing and metagenomics is an important tool for diagnosing viral infections and managing treatment

Over the last years, many countries have incorporated next generation sequencing (NGS) technologies to integrate pathogen genome sequencing into infectious disease surveillance. NGS can also be a powerful diagnostic tool for pathogen detection. Currently, different approaches that include deep sequencing of genomic and sub-genomic viral segments, targeted metagenomics, and shotgun sequencing are all being explored in the clinic, and to characterize the human virome. Viral discovery is also

advancing at a rapid pace in complementary areas such as veterinary medicine, where NGS can improve detection of novel emerging pathogens and herd screening (1, 2).

Perhaps the most interesting application of NGS for viral discovery lies in the identification of etiologic agents in individuals with complex clinical conditions and suspected infections where pathogen recovery with conventional tests have failed. The shotgun approach is the most comprehensive but also the most challenging, as it allows testing for any possible pathogen in the sample with no a priori- assumptions. The analysis includes removal of human and non-relevant sequences to enrich the viral sequences that might be present in the sample at a level as low as 1%. This approach proves to be powerful for broad-range detection of known and emerging viruses, including early detection of zoonotic infections, which are not included in the most comprehensive multiplex real-time PCR panels such as the FilmArray® [reviewed in (3)]. Besides viral discovery, metagenomics-based testing also enables genotyping, assessment of molecular markers for drug resistance, and molecular epidemiologic studies. Some have also speculated that NGS can be used to quantify viral loads, as reads per kilobase of reference sequence per million total sequencing reads have good correlation with real time quantitative PCR (qPCR) viral copies per milliliter for respiratory viruses (*P* value <0.0001 across viral taxa) (4).

Because of its potential advantage to detect any possible pathogen (bacteria, viruses, fungi, parasites, rare or even unknown pathogens) in a single test and with high sensitivity, a number of studies have evaluated the clinical utility of NGS shotgun metagenomics sequencing for diagnosis of viral infections. In patients with suspected infections of the central nervous system, multiple studies agree that NGS yields higher positive pathogen detection rates and detect more viruses at a faster time in comparison to conventional viral diagnostic tests (5, 6). However, NGS seems to be less sensitive compared to the standard amplification-based assays in the diagnosis of encephalitis, where low viral loads are common. When a validated NGS protocol was applied to RT-PCR/PCR positive cerebrospinal fluid (CSF) samples from patients with low viral load known viral encephalitis, NGS failed to detect the viruses in 71.4% of the cases (7). In immunocompromised patients, NGS allowed identification of pathogens (which were mostly viruses) at a higher rate of detection and at a significantly faster time than using standard methods (8). Also, NGS proved to be a strong method as first-line diagnostic tool in individuals with compromised immunity, as it identified three times more potentially relevant viruses in samples obtained at inclusion than conventional microbiological methods (36% vs. 11%) (9). In patients with connective tissue diseases, however, there is still debate as to what role opportunistic viruses such as CMV and EBV play in the pathogenesis of the diseases (10, 11). Therefore, while NGS has the advantage of being able to identify a wide range of potential pathogens, the importance of genomic tests in implementing antiviral therapy or in guiding clinical management is still limited to the availability of antiviral drugs and to a better understanding of what role viruses play in complex scenarios.

A less comprehensive, but still very powerful NGS approach for the identification of many viral infections use different methods to select for or enrich the sample in sequences of infectious agents known to cause infections in humans and/or animals. This targeted metagenomics may be more easily adapted as a sequence-based diagnostic assay, with the limitations that it will only detect specific pathogens and that it will require additional validations in case a new pathogen is discovered. In a recent study, Castellot and colleagues used a pan viral (DNA and RNA viruses) metagenomics approach to study cerebrospinal fluid (CSF) samples from 40 pediatric patients with meningoencephalitis of unknown etiology. The test yielded twenty additional putative virological diagnoses to the seven previously obtained through multiplex real-time PCR for HHV1-2, VZV and enterovirus routinely used for diagnosis of viral meningoencephalitis, although only 10 of the 23 viruses were confirmed by specific PCR (12).

The use of genomic-based tests for managing HIV-1 antiretroviral therapy

Besides being a powerful tool for viral discovery, NGS provides high-resolution characterization of individual mutations in the viral genomes. Some of these mutations have a phenotypic or biological correlate affecting disease progression or response to therapy, and therefore may provide insights into drug susceptibility and treatment strategy. In this regard, the HIV-1 epidemic stands as a significant example of the pivotal role of viral genomics in advancing healthcare and improving clinical decisions.

Because HIV can mutate rapidly and develop resistance to almost any anti-HIV drug, it became evident that testing for drug resistance mutations (DRMs) could help guide the choice of antiretroviral drugs (ARVs) for combination antiretroviral therapy (ART), improve virological response, and maintain suppression in people living with HIV (PLWH). Indeed, since its introduction in the January 2000 Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents by the DHHS (13), testing for HIV drug resistance (HIVDR) became part of the standard of care in PLWH.

Beyond technical and methodological breakthroughs, a large number of studies support the fundamental correlations that underlie HIVDR knowledge and set the scientific basis for genotype-based HIVDR tests. Namely, genotype-treatment, genotype-phenotype, and genotype-outcome correlations. These studies, together with prospective and retrospective data collected from clinical trials and cohort studies were key to understanding the mechanisms and evolution of HIVDR in real-world settings, and became significant for pharmaceutical companies to produce new ARVs that could be used in PLWH that develop resistance mutations to previously available ARVs of the same class. By 2007, FDA Guidance for Industry recommended characterization of resistance and cross-resistance during ARV drug development and in the post-marketing setting, upholding HIVDR testing (14). Having the genomic information at hand, and matching it with that of structural data and computational methods for measuring enzyme structure and protein-drug interactions accelerated discoveries that were used to develop compounds that inhibit both the wild-type and the drug resistant mutant viral enzymes.

An example of this was the *in silico* development of dolutegravir, a highly potent and effective second generation integrase strand transfer inhibitor (INSTI) that is currently being used for treatment of PLWH worldwide.

Interpretation of HIV-1 genotypic resistance tests is complex because there are many different drug resistance associated mutations, which occur in complex patterns and which have diverse effects on the ARVs within each drug class. Thus, interpretation of HIVDR often requires consultation with an expert in HIV drug resistance. Different bioinformatic tools have been developed to aid virologists and clinicians in the interpretation of HIV-1 genotypic resistance, such as the freely available web-based Stanford HIVdb program (https://hivdb.stanford.edu/hivdb/). This tool works by aligning the sub-genomic HIV-1 sequences to a reference HIV-1 genome, and then comparing the mutations to a list of pre-defined drug resistance associated mutations, finally informing the level of drug resistance to each ARV. Therefore, the accuracy of the interpretation system depends on regular updates based on a thorough review of data from sequence databases such as GenBank, peer-reviewed publications, and often from the authors of these publications.

New ARVs coming down the pipeline or entering clinical use for HIV target regions of the genome that include not only protease (PR) and reverse transcriptase (RT), but also other regions. For example, integrase inhibitors target integrase (INT); CCR5 inhibitors target envelope gp120, fusion inhibitors, and monoclonal antibodies target envelope gp41, and capsid inhibitors target gag capsid (CA). Thus, NGS provides a better option than Sanger sequencing, as it has the advantage of sequencing several genomic segments at the same time. Also, NGS can inform the presence of minority mutations (those present in 1% to 20% of the viral population), which may increase the risk of treatment failure [15–17]. This additional information regarding relative abundance of susceptible/resistant strains strengthens our ability to assess the clinical impact of a given DRM and to determine and track its overall frequency within a population, which may significantly impact drug regimens and public health approaches to control and reduce HIV [18]. Besides its clinically validated use of informing HIVDR, NGS has the potential to distinguish between defective and potentially replication-competent HIV-1 proviruses [19]; provide phylogenetic relationships, allow for identification of novel HIV-1 genotypes [20], and even inform HIV viral loads [21]. Finally, the emergence of long read third generation sequencing of single viral genomes may be of future value for identifying linked mutations and multi drug resistant HIV.

Challenges to adopting NGS and metagenomics for diagnostics and treatment of viral infections

NGS workflows are more intricate compared to traditional diagnostic methods because they involve multiple steps, each with potential variables that need to be controlled. The inherent complexity of NGS requires the validation of a wide range of equipment, methods, and processes to ensure accurate, reproducible, and reliable diagnostic results. Rigorous assessment of the assay's reliability should also involve reproducibility across testing personnel and repeated runs, which is costly and difficult to implement. In addition, findings usually need confirmation by a second validated method, making NGS difficult to incorporate in clinical settings. Preparation of libraries involve a number of steps that differ widely depending on the protocol involved. If sample transport is required, pre-analytical parameters will also need to be standardized. It is important to consider sampling of the appropriate type of tissue or source, and at the appropriate time of sampling, not to mention the interpretation of contaminant infectious agents of no clinical relevance.

NGS library preparation kits and software that can identify a panel or group of pre-defined pathogens, or determine the antiviral resistance profile of a virus based on sequence capture-based partial or whole genome sequencing are easily validated through comparison with Sanger sequencing [22, 23]. However, performance testing for viral metagenomics for pathogen detection is intrinsically different from benchmarking of tests based on targeted approaches. At present, there are no standard procedures or parameters inherent to clinical diagnostic use that can correctly reflect quality, such as specificity calculations, sensitivity for divergent viruses and variants, and importantly, a determined cut-off for defining a positive result for each workflow for virome analyses. Thus, the sharing, comparison, and reliable production of the results of analyses are difficult. The increasing number of bioinformatics tools for analysis and the dynamic characteristic of databases is another challenge. Using an artificial simulated in silico data set of >6 million single-end 150-bp Illumina HiSeq sequences derived from viral genomes, human chromosomes, and bacterial DNA, Brinkmann et al. evaluated how different levels of experience and/or bioinformatics methodologies affect the outputs and interpretation of viral pathogens in high throughput sequence data [24]. In this study, 13 different European institutes for bioinformatics analysis were allowed to use their bioinformatics tools and workflows of choice to evaluate sensitivity, specificity, turnaround time and interpretation of results. Results showed difficulties in the identification of mutated and divergent viral sequences, and selective effects probably due to the choice of different viral databases. This was also observed in a benchmark of 13 metagenomic pipelines currently used in clinical virological laboratories that adhered to the European Society for Clinical Virology Network recommendations for wet-lab [25] and bioinformatics analysis and reporting [26]. In this study, positive predictive value (PPV) varied from 71% to 100%, and read counts of target viruses over a range of 2-3 log, indicating difficulties in detection of low abundant viral pathogens and mixed infections for clinical diagnostic use. Perhaps the most challenging step of diagnostic NGS tests is the in silico bioinformatics analysis of the NGS data known as the postanalytics. While difficult to harmonize, this step is crucial for reporting, and involves interpretation by an experienced, qualified health professional to correlate bioinformatics results with clinical and epidemiological patient information. A recent evaluation of the clinical impact of a standardized NGS plasma cell-free DNA (Karius test) in a retrospective multicenter cohort study showed that even when all the variables are controlled and the test is performed in an accredited laboratory, diagnostic interpretation algorithms are missing to define the real-world impact

of these tests in clinical practice (27). How to standardize the definition and application of clinical impact criteria in complex medical conditions to ensure consistent evidence-based care will definitively need to be approached in clinical virology.

Conclusions

NGS allows for identification not only for pathogens previously considered in clinicians' diagnosis algorithms, but also pathogens for which corresponding diagnostic tests are unavailable in clinical microbiology laboratories or ineffective because of genetic divergence of the infectious agent. It could also identify co-infections that exacerbate or drive disease. In addition to detection, the same test could provide a quantification of the pathogen load, and genotypic markers of drug resistance and pathogenicity. NGS has shown to impact positively on time to diagnosis, tailored therapeutics and improved overall survival of patients with diverse medical conditions. However, several challenges will need to be overcome in order to expand its use in clinical virology beyond viral surveillance. First, the cost is still high, with no guarantee of a result. Validation and standardization of methods and analysis tools are needed, as well as better trained multidisciplinary teams involving bioinformaticians, clinical virologists and disease specialists that can interpret the results and maximize the benefit.

Data availability statement

The datasets presented in this article are not readily available because the perspective provides original theory and opinion from the authors. Requests to access the datasets should be directed to PA, pauauli@gmail.com.

Author contributions

PA: Writing – original draft. JK: Conceptualization, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Chen J, Suo X, Cao L, Yuan C, Shi L, Duan Y, et al. Virome analysis for identification of a novel porcine sapelovirus isolated in Western China. *Microbiol Spectr.* (2022) 10:e0180122. doi: 10.1128/spectrum.01801-22

2. Kubacki J, Fraefel C, Bachofen C. Implementation of next-generation sequencing for virus identification in veterinary diagnostic laboratories. *J Vet Diagn Invest.* (2021) 33:235–47. doi: 10.1177/1040638720982630

3. Wani AK, Chopra C, Dhanjal DS, Akhtar N, Singh H, Bhau P, et al. Metagenomics in the fight against zoonotic viral infections: A focus on SARS-CoV-2 analogues. *J Virol Methods.* (2024) 323:114837. doi: 10.1016/j.jviromet.2023.114837

4. Graf EH, Simmon KE, Tardif KD, Hymas W, Flygare S, Eilbeck K, et al. Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: a systematic comparison to a commercial PCR panel. *J Clin Microbiol.* (2016) 54:1000–7. doi: 10.1128/JCM.03060-15

5. Yuan L, Zhu XY, Lai LM, Chen Q, Liu Y, Zhao R. Clinical application and evaluation of metagenomic next-generation sequencing in pathogen detection for suspected central nervous system infections. *Sci Rep.* (2024) 14:16961. doi: 10.1038/s41598-024-68034-1

6. Zhang S, Wu G, Shi Y, Liu T, Xu L, Dai Y, et al. Understanding etiology of community-acquired central nervous system infections using metagenomic next-generation sequencing. *Front Cell Infect Microbiol.* (2022) 12:979086. doi: 10.3389/fcimb.2022.979086

7. Perlejewski K, Bukowska-Ośko I, Rydzanicz M, Pawełczyk A, Caraballo Cortés K, Osuch S, et al. Next-generation sequencing in the diagnosis of viral encephalitis: sensitivity and clinical limitations. *Sci Rep.* (2020) 10:16173. doi: 10.1038/s41598-020-73156-3

8. Tang W, Zhang Y, Luo C, Zhou L, Zhang Z, Tang X, et al. Clinical application of metagenomic next-generation sequencing for suspected infections in patients with primary immunodeficiency disease. *Front Immunol.* (2021) 12:696403. doi: 10.3389/fimmu.2021.696403

9. Parize P, Pilmis B, Lanternier F, Lortholary O, Lecuit M, Muth E, et al. Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study. *Clin Microbiol Infection.* (2017) 23:574.e1–6. doi: 10.1016/j.cmi.2017.02.006

10. Su R, Yan H, Li N, Ding T, Li B, Xie Y, et al. Application value of blood metagenomic next-generation sequencing in patients with connective tissue diseases. *Front Immunol.* (2022) 13:939057. doi: 10.3389/fimmu.2022.939057

11. Wang H, Shi X, Yang H, Du Y, Xue J. Metagenomic next-generation sequencing shotgun for the diagnosis of infection in connective tissue diseases: A retrospective study. *Front Cell Infect Microbiol.* (2022) 12:865637. doi: 10.3389/fcimb.2022.865637

12. Castellot A, Camacho J, Fernández-García MD, Tarragó D. Shotgun metagenomics to investigate unknown viral etiologies of pediatric meningoencephalitis. *PloS One.* (2023) 18:e0296036. doi: 10.1371/journal.pone.0296036

13. U.S. Department of health and human services (DHHS) guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents(2000). Available online at: https://clinicalinfo.hiv.gov/sites/default/files/guidelines/archive/AdultandAdolescentGL02052001009.pdf. (Accessed October 01, 2024)

14. U.S. Department of Health and Human Services. *Food and Drug Administration. Center for Drug Evaluation and Research. Guidance for Industry - Role of HIV Resistance Testing in Antiretroviral Drug Development* (2007). Available online at: http://www.fda.gov/cder/guidance/index.htm. (Accessed October 01, 2024)

15. Paredes R, Lalama CM, Ribaudo HJ, Schackman BR, Shikuma C, Gigue F, et al. Pre-existing minority drug-resistant HIV-1 variants, adherence, and risk of antiretroviral treatment failure. *J Infect Dis.* (2010) 201:662–71. doi: 10.1086/650543

16. Simen BB, Simons JF, Hullsiek KH, Novak RM, MacArthur RD, Baxter JD, et al. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J Infect Dis*. (2009) 199:693–701. doi: 10.1086/596736

17. Johnson JA, Li JF, Wei X, Lipscomb J, Irlbeck D, Craig C, et al. Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy. *PloS Med*. (2008) 5:e158. doi: 10.1371/journal.pmed.0050158

18. Lee ER, Parkin N, Jennings C, Brumme CJ, Enns E, Casadellà M, et al. Performance comparison of next generation sequencing analysis pipelines for HIV-1 drug resistance testing. *Sci Rep*. (2020) 10:1634. doi: 10.1038/s41598-020-58544-z

19. Hiener B, Eden JS, Horsburgh BA, Palmer S. Amplification of near full-length HIV-1 proviruses for next-generation sequencing. *J Vis Exp*. (2018) 140):58016. doi: 10.3791/58016

20. Yamaguchi J, Vallari A, McArthur C, Sthreshley L, Cloherty GA, Berg MG, et al. Brief report: complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. *J Acquir Immune Defic Syndr*. (2020) 83:319–22. doi: 10.1097/QAI.0000000000002246

21. Fogel JM, Bonsall D, Cummings V, Bowden R, Golubchik T, De Cesare M, et al. Performance of a high-throughput next-generation sequencing method for analysis of HIV drug resistance and viral load. *J Antimicrob Chemother*. (2020) 75:3510–6. doi: 10.1093/jac/dkaa352

22. Jenkins F, Le T, Farhat R, Pinto A, Anzari A, Bonsall D, et al. Validation of an HIV whole genome sequencing method for HIV drug resistance testing in an Australian clinical microbiology laboratory. *J Med Virol*. (2023) 95:e29273. doi: 10.1002/jmv.29273

23. Manso CF, Bibby DF, Lythgow K, Mohamed H, Myers R, Williams D, et al. Technical validation of a hepatitis C virus whole genome sequencing assay for detection of genotype and antiviral resistance in the clinical pathway. *Front Microbiol*. (2020) 11:576572. doi: 10.3389/fmicb.2020.576572

24. Brinkmann A, Andrusch A, Belka A, Wylezich C, Höper D, Pohlmann A, et al. Proficiency testing of virus diagnostics based on bioinformatics analysis of simulated in silico high-throughput sequencing data sets. *J Clin Microbiol*. (2019) 57:e00466–19. doi: 10.1128/JCM.00466-19

25. López-Labrador FX, Brown JR, Fischer N, Harvala H, Van Boheemen S, Cinek O, et al. Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: Wet lab procedure. *J Clin Virol*. (2021) 134:104691. doi: 10.1016/j.jcv.2020.104691

26. de Vries JJC, Brown JR, Couto N, Beer M, Le Mercier P, Sidorov I, et al. Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: bioinformatic analysis and reporting. *J Clin Virol*. (2021) 138:104812. doi: 10.1016/j.jcv.2021.104812

27. Hogan CA, Yang S, Garner OB, Green DA, Gomez CA, DIen Bard J, et al. Clinical impact of metagenomic next-generation sequencing of plasma cell-free DNA for the diagnosis of infectious diseases: A multicenter retrospective cohort study. *Clin Infect Dis*. (2021) 72:239–45. doi: 10.1093/cid/ciaa035