The Emergence of Pathogen Genomics in Diagnostic Laboratories

ERIN GRAF

LEARNING OBJECTIVES

1. Describe the basics of next-generation sequencing (NGS) workflows for clinical microbiology applications.
2. Identify barriers to widespread adoption of NGS in clinical microbiology laboratories.
3. Define workforce needs associated with NGS applications in clinical microbiology.

ABSTRACT

Next-generation sequencing (NGS) technologies enable one to obtain genomic information about pathogens directly from clinical samples or isolates on a scale never before possible with polymerase chain reaction or Sanger sequencing. Clinical microbiology laboratories are rapidly finding the "low-hanging fruit"–type applications for this disruptive technology as it becomes cost-effective with reasonable turn-around times. This includes the use of metagenomic NGS for pathogen detection in primary clinical samples. It also includes the use of NGS for detection of antimicrobial-resistance genes in bacteria that are difficult and/or slow to grow in culture. Finally, NGS offers improvement in resolution of viral and bacterial outbreaks and streamlines workflows for a "one-size-fits-all" protocol. There are still significant barriers to implementation that must be considered. These include workforce skills required for high-complexity wet-bench protocols and data analysis, as well as validation approaches for interpretation given the lack of US Food and Drug Administration–approved systems. Regardless, NGS will continue to be a game changer in clinical microbiology, and laboratories must be prepared to face the challenges associated with adoption of this technology.

ABBREVIATIONS: CLS - clinical laboratory scientist, FDA - US Food and Drug Administration, mNGS - metagenomic NGS, NGS - next-generation sequencing, WGS - whole genome sequencing.

INDEX TERMS: next-generation sequencing, metagenomics, whole genome sequencing.

***Erin Graf**, Children's Hospital of Philadelphia*

***Address for Correspondence: Erin Graf**, Children's Hospital of Philadelphia, erinhgraf@gmail.com*

Clin Lab Sci 2019;32(2):67–69

INTRODUCTION

Technologic advances in sequencing have revolutionized molecular pathology. Clinical human genomics laboratories have used next-generation sequencing (NGS) approaches for the diagnosis of germline and somatic disorders for nearly a decade.[1] Clinical microbiology laboratories have been slower to apply NGS for routine clinical use because of the cost and speed and because, at present, only a small number of case reports show a benefit to patient care.[2] As a result, most studies have been limited to research settings and a handful of reference laboratories.[3-7] This will no doubt change as clinical microbiology laboratories become more familiar with the technology and the advantages it may provide over conventional methods. Furthermore, as sequencing platforms continue to decrease in price and turnaround times, laboratories may consider replacement of current or conventional methods to maximize cost-effectiveness.

NGS WORKFLOWS IN CLINICAL MICROBIOLOGY

There are 3 primary workflow approaches to use of NGS in clinical microbiology. The first is termed shotgun metagenomics, often referred to as metagenomic NGS (mNGS).[2] In this approach, all nucleic acid (RNA or DNA) in a clinical sample is sequenced without any selection. Several million sequences are generated and analyzed to look for pathogens in a high background of human nucleic acid.[8] An alternative approach, which requires less total sequence for analysis and is therefore cheaper, is amplicon NGS. This is the basis for many human microbiome studies, wherein conserved regions common to all bacteria are amplified.[9] Finally, an isolate grown in culture can also be sequenced at a cheaper cost compared with mNGS. This is generally termed whole genome sequencing (WGS),[10] though a pathogen whole genome sequence can be generated via mNGS.

The basic steps involved in performing NGS are generally similar regardless of the application above. For the most commonly used short-read sequencing platforms, nucleic acid (RNA or DNA) is chopped up to allow for smaller fragments to be sequenced in parallel. These fragments are ligated to identical small sequences that act as (1) binding agents to fix the fragments to the sequencing

Downloaded from http://hwmaint.clsjournal.ascls.org/ on September 14 2026

chip or bead and (2) primer sequences that allow the sequencing reactions to happen.[11] Barcodes can also be added to create unique sample identities to allow for sample pooling and, therefore, cost-effective sequencing. The process of library preparation can take several hours to days, with numerous hands-on steps, depending on the protocol used. Automation is possible but may not be cost effective for low-volume applications and does not necessarily result in reduced turnaround times.

After quality-control steps, whereby library concentration and size are calculated for user-defined pooling ratios, libraries are loaded onto the sequencer. The time from loading to data extraction is dependent upon the instrument used and desired sequence length. For most clinical microbiology-related applications, the fastest run time is around 18 hours. Once available, data are extracted in the form of fastq files. These are lines of nucleotide sequence, or reads, with accompanying quality scores. The size of each fastq file is dependent upon the size of the bead or chip used and the amount of space occupied by a sample (Ie, pooling ratio). The average file size for metagenomic data, which requires several million reads, is on the order of gigabytes, and an entire sequencing run can be on the order of terabytes.

Given the file sizes above, computational and storage needs are important considerations. Data analysis requires at least some basic bioinformatics knowledge to use commercially available pipelines, even those with graphical user interfaces.[12] Many analysis pipelines require knowledge of command line interface, and thus experience with programming is critical. Most clinical laboratories that perform NGS-based testing have a separate team of bioinformaticians and programmers that interface with the clinical laboratory to validate, streamline, and perform routine analyses. Depending on the applications to clinical microbiology and pipeline(s) used, generation of interpretable results can take minutes to several hours.[13,14] The final step in the process is a director-level review and sign-out of results in a clinician-friendly report.

APPLICATIONS OF PATHOGEN GENOMICS IN CLINICAL MICROBIOLOGY

As NGS technologies continue to decrease in cost and turnaround time, clinical microbiology laboratories have considered unmet needs that deeper sequencing can fill. This includes detection of pathogens in cases in which conventional microbiology testing has been unable to find a cause. To date, only a handful of reference laboratories offer this as an orderable test, and each has focused on a particular specimen type (ie, cerebrospinal fluid, plasma, bronchoalveolar lavage fluid) for validation.[8,15,16] A second unmet need is faster antimicrobial susceptibility testing to support optimal antimicrobial stewardship. For some pathogens, there can be a delay of days to weeks before a physician knows whether the empiric treatment prescribed is effective or whether therapy can be narrowed for stewardship purposes. With NGS, it is possible this window could be shortened for some pathogens that grow very slowly or do not grow in culture.[17] Finally, understanding transmission dynamics of infectious pathogens at the hospital level all the way to the global level is important for infection control and public health programs. The current methods available to determine pathogen relatedness provide low resolution and are highly pathogen specific, limiting widespread adoption.[18,19] Furthermore, as we continue to replace culture with molecular panels, isolates will no longer be available for these conventional approaches.[20] NGS can overcome these limitations, providing higher resolution for strain comparisons and allowing for genomic analyses directly from a clinical sample without the need for culture in a single protocol.

BARRIERS TO ROUTINE CLINICAL USE OF NGS

As NGS takes hold in clinical microbiology laboratories, it is important to consider the workforce needs for successful implementation. Many clinical laboratory scientists (CLSs) do not have formal training on the wet bench and, even less likely, the analysis sides of NGS for human genomics, let alone any microbiology applications. These are skills that will need to be gained during on-the-job training without more formal curricula embedded in CLS training programs.[21]

The data derived from any of the applications above are of the highest complexity encountered in any laboratory. This is due, in part, to the difficulty in determining clinical significance. For example, there are no set standards for the number of reads required to report a positive detection of a pathogen from mNGS.[8] Pathogen reads must be distinguished from both contamination and commensal flora, a problem shared by conventional microbiologic methods (eg, culture). Similarly, the database of resistance genes continues to evolve and not every mutation associated with antimicrobial resistance is known, limiting the ability to call pathogens "susceptible" to a particular drug.[17] Finally, there are no US Food and Drug Administration (FDA)–approved methods for these applications. Thus, the burden falls on laboratories to validate their own methods and criteria for reporting. The College of American Pathologists provides some limited validation guidance via their Molecular Pathology checklist. However, FDA-approved sample-to-answer solutions, analogous to multiplex polymerase chain reaction panels, may be needed for widespread adoption of NGS in clinical microbiology laboratories.

SERIES FOCUS

This series will provide readers with detailed understanding of the 3 primary applications of pathogen genomics in the diagnostic setting. Each review will discuss the

Downloaded from http://hwmaint.clsjournal.ascls.org/ on September 14 2026

limitations of the specific approach and highlight some of the equipment and/or workforce needs for successful implementation. The second article in this series, "Use of Diagnostic Metagenomics in the Clinical Microbiology Laboratory," will discuss mNGS for detection of pathogens directly from clinical samples. As mentioned above, reports of successful real-time clinical use of mNGS are quite scarce, and larger prospective studies are needed to show the benefit of adoption. The third article, "Next Generation Sequencing for Outbreak Investigation in the Clinical Microbiology Laboratory," reviews the advances in understanding pathogen transmission dynamics by performing higher-resolution WGS via NGS. The final article, "Use of whole genome sequencing for detection of antimicrobial resistance: *Mycobacterium tuberculosis*, a model organism," describes the challenges with inferring phenotype from genotype via NGS. It also highlights the most likely replacement of culture-based susceptibility testing with WGS, that is, specifically for the pathogen *Mycobacterium tuberculosis*. The series shares common themes of current barriers to widespread adoption of any NGS application in clinical microbiology laboratories.

REFERENCES

1. Müllauer L. Milestones in pathology-from histology to molecular biology. *Mag Eur Med Oncol*. 2017;10(1):42–45. doi: 10.1007/s12254-016-0307-z
2. Simner PJ, Miller S, Carroll KC. Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin Infect Dis*. 2018;66(5):778–788. doi: 10.1093/cid/cix881
3. Abril MK, Barnett AS, Wegermann K, et al. Diagnosis of *Capnocytophaga canimorsus* sepsis by whole-genome next-generation sequencing. *Open Forum Infect Dis*. 2016;3 (3):ofw144. doi: 10.1093/ofid/ofw144
4. Hong DK, Blauwkamp TA, Kertesz M, Bercovici S, Truong C, Banaei N. Liquid biopsy for infectious diseases: sequencing of cell-free plasma to detect pathogen DNA in patients with invasive fungal disease. *Diagn Microbiol Infect Dis*. 2018;92 (3):210–213. doi: 10.1016/j.diagmicrobio.2018.06.009
5. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med*. 2014;370(25):2408–2417. doi: 10.1056/NEJMoa1401268
6. Chiu CY, Coffey LL, Murkey J, et al. Diagnosis of fatal human case of St. Louis encephalitis virus infection by metagenomic sequencing, California, 2016. *Emerg Infect Dis*. 2017;23 (10):1964–1968. doi: 10.3201/eid2310.161986
7. Murkey JA, Chew KW, Carlson M, et al. Hepatitis E virus-associated meningoencephalitis in a lung transplant recipient diagnosed by clinical metagenomic sequencing. *Open Forum Infect Dis*. 2017;4(3):ofx121. doi: 10.1093/ofid/ofx121
8. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G; Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology; Microbiology Resource Committee of the College of American Pathologists. Validation of Metagenomic next-generation sequencing tests for universal pathogen detection. *Arch Pathol Lab Med*. 2017;141(6):776–786. doi: 10.5858/arpa.2016-0539-RA
9. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486 (7402):222–227. doi: 10.1038/nature11053
10. Gilchrist CA, Turner SD, Riley MF, Petri WA Jr, Hewlett EL. Whole-genome sequencing in outbreak analysis. *Clin Microbiol Rev*. 2015;28(3):541–63. doi: 10.1128/CMR.00075-13
11. Buermans HP, den Dunnen JT. Next generation sequencing technology: advances and applications. *Biochim Biophys Acta*. 2014;1842(10):1932–1941. doi: 10.1016/j.bbadis.2014.06.015
12. Carriço JA, Rossi M, Moran-Gilad J, Van Domselaar G, Ramirez M. A primer on microbial bioinformatics for non-bioinformaticians. *Clin Microbiol Infect*. 2018;24(4):342–349. doi: 10.1016/j.cmi.2017.12.015
13. Flygare S, Simmon K, Miller C, et al. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol*. 2016;17(1):111. doi: 10.1186/s13059-016-0969-1
14. Naccache SN, Federman S, Veeraraghavan N, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res*. 2014;24(7):1180–1192. doi: 10.1101/gr.171934.113
15. Blauwkamp TA, Thair S, Rosen M, et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol*. 2019;4(4):663–674. doi: 10.1038/s41564-018-0349-6
16. Miller S, Naccache SN, Samayoa E, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res*. 2019;29 (5):831–842. doi: 10.1101/gr.238170.118
17. Ellington MJ, Ekelund O, Aarestrup FM, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin Microbiol Infect*. 2017;23(1):2–22. doi: 10.1016/j.cmi.2016.11.012
18. Herschleb J, Ananiev G, Schwartz DC. Pulsed-field gel electrophoresis. *Nat Protoc*. 2007;2(3):677–684. doi: 10.1038/nprot.2007.94
19. Maiden MC, Jansen van Rensburg MJ, Bray JE, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol*. 2013;11(10):728–736. doi: 10.1038/nrmicro3093
20. Binnicker MJ, Multiplex molecular panels for diagnosis of gastrointestinal infection: performance, result interpretation, and cost-effectiveness. *J Clin Microbiol*. 2015;53(12):3723–3728. doi: 10.1128/JCM.02103-15
21. Taylor S, Bennett KM, Deignan JL et al. Molecular pathology curriculum for medical laboratory scientists: a report of the association for molecular pathology training and education committee. *J Mol Diagn*. 2014;16(3):288–296. doi: 10.1016/j.jmoldx.2014.02.003

Downloaded from http://hwmaint.clsjournal.ascls.org/ on September 14 2026