The Journal of Molecular Diagnostics, Vol. ■, No. ■, ■ 2016

the Journal of Molecular Diagnostics
jmd.amjpathol.org

MINI-REVIEW

In Silico Proficiency Testing for Clinical Next-Generation Sequencing

Q8 Eric J. Duncavage,* Haley J. Abel,† and John D. Pfeifer*

From the Departments of Pathology and Genetics,† Washington University School of Medicine, St. Louis, Missouri*

Accepted for publication September 13, 2016.

Address correspondence to John D. Pfeifer, M.D., Ph.D., Department of Pathology, Washington University School of Medicine, Campus Box 8118, 660 S Euclid Ave, St. Louis, MO 63110. E-mail: pfeifer@path.wusltl.edu.

Quality assurance for clinical next-generation sequencing (NGS)—based assays is difficult given the complex methods and the range of sequence variants such assays can detect. As the number and range of mutations detected by clinical NGS assays has increased, it is difficult to apply standard analyte-specific proficiency testing (PT). Most current proficiency testing challenges for NGS are methods-based PT surveys that use DNA from reference samples engineered to harbor specific mutations that test both sequence generation and bioinformatics analysis. These methods-based PTs are limited by the number and types of mutations that can be physically introduced into a single DNA sample. *In silico* proficiency testing, which evaluates only the bioinformatics component of NGS assays, is a recently introduced PT method that allows for evaluation of numerous mutations spanning a range of variant classes. *In silico* PT data sets can be generated from simulated or actual sequencing data and are used to test alignment through variant detection and annotation steps. *In silico* PT has several advantages over the use of physical samples, including greater flexibility in tested variants, the ability to design laboratory-specific challenges, and lower costs. Herein, we review the use of *in silico* PT as an alternative to traditional methods-based PT as it is evolving in oncology applications and discuss how the approach is applicable more broadly. *(J Mol Diagn 2016, ■: 1—8; http://dx.doi.org/10.1016/j.jmoldx.2016.09.005)*

Q3 Massively parallel sequencing technologies [alias next-generation sequencing (NGS) technologies] have a level of technical complexity unmatched by most clinical laboratory testing methods.[1,2] This complexity can be modeled in three distinct components, specifically, the wet bench processes involved in DNA target enrichment and library preparation; the sequencing platforms responsible for generation of raw sequence reads; and the bioinformatics pipelines (often called the dry bench component of NGS) that are used to identify variants within the sequence reads. For clinical NGS assays, each of these three components needs to be optimized, individually as well as with respect to one another, depending on the size of the target region (eg, the number of genes in a panel versus the exome versus the whole genome), assay design (eg, amplification versus hybrid capture-based), disease setting (eg, testing for constitutional or inherited variants versus testing of cancer specimens for acquired somatic variants), and types of variants expected (single-nucleotide variants versus insertions, deletions, and structural variants).[3]

In addition to the intricacies of the analytic phase of NGS, the range of variants and variant types that can be evaluated by NGS methods within a single assay is also unmatched. Optimized NGS assays can detect a full range of variant types, including single-nucleotide variants; small insertions and deletions; copy number variants; and structural variants, such as translocations and inversions.[4–9] NGS assays make it possible to detect not only the presence of sequence variants, but also the variant allele frequency (VAF; sequence reads containing the variant/total reads that contain the individual base position). Given the range of variant types and frequencies that can be detected, and the different assay designs, library preparation approaches, sequencing platforms, and bioinformatics pipelines, the

Supported by Washington University School of Medicine departmental funding.

Q1 Q2 Disclosures: E.J.D., H.J.A., and J.D.P. are cofounders of P&V Licensing LLC. P&V Licensing LLC is a company that provides *in silico* mutagenesis of next-generation sequencing sequence files.

http://dx.doi.org/10.1016/j.jmoldx.2016.09.005

Table 1 Summary of Existing Approaches to EQA and PT for NGS

Approach	Uses	Advantages	Limitations
EQA PT (eg, distribution of highly purified nucleic acids extracted from human cell lines provided by a CLIA 1988 accepted provider)	Conventional analyte-based PT	Closely replicates clinical samples; can assess performance of entire test system	Only available for a small fraction of NGS tests used in routine clinical practice
Alternative performance assessment (eg, split sample analysis with other laboratories or in-house)	Conventional analyte-based PT; implemented when an EQA survey by a CLIA 1988 accepted provider is not available	Well-established alternative approach to PT; closely replicates clinical samples; can assess performance of entire test system	Requires time-consuming individualized design and implementation for each analyte
Methods-based proficiency testing (eg, based on the method rather than each individual analyte)	Comparison of important aspects of analysis and interpretation; implemented when there are no (or few) analytes in common between laboratories	Well-established paradigm; supported by CMS; can be wet bench or dry bench based*; CAP provides several MBPT surveys	Must be furnished by a CLIA 1988 accepted provider; focused on a step or process within the test system (ie, does not assess performance of entire test system)

Information herein is condensed from Kalman et al[11] and Schrijver et al.[12]
*See Table 2.
CAP, College of American Pathologists; CLIA, Clinical Laboratory Improvement Amendment; CMS, Center for Medicare and Medicaid Services; EQA, external quality assessment; MBPT, methods-based proficiency testing; NGS, next-generation sequencing; PT, proficiency testing.

challenges to comprehensive quality control and quality assurance are unprecedented.

Existing Approaches to External Quality Assessment and Proficiency Testing

As with all clinical laboratory tests, NGS assays must undergo external quality assessment, as mandated by Clinical Laboratory Improvement Amendment 1988,[10,11] and proficiency testing (PT) is one of the primary external measures of quality. The range of assay designs, target regions, variants types and VAFs, library preparation approaches, sequencing platforms, and bioinformatics pipelines make the analyte-specific PT programs that are the traditional approach to external quality assessment insufficient for NGS tests (Table 1). In contrast, methods-based proficiency testing (MBPT) methods are ideally suited to NGS assays.[11–13]

Comprehensive MBPT methods are based on biological specimens, and include neoplastic cell lines directly derived from human neoplasms and cell lines that have been engineered to harbor specific sequence variants. Cell lines have the advantage that they are renewable reagents that can be mixed at known ratios to simulate different VAFs. And because formalin-fixed, paraffin-embedded (FPPE) cell blocks can easily be produced from cell lines, cell lines are useful sources for MBPT in the oncology setting in that they can be used to simultaneously evaluate most aspects of the workflow of an NGS test (eg, as a fresh cell pellet for workflows that involve peripheral blood, or as a formalin-fixed, paraffin-embedded cell block for workflows that involve testing of formalin-fixed, paraffin-embedded solid tumor specimens). Several governmental and commercial entities have developed well-characterized cell lines for sequence variants specific to several genetic conditions; the Centers for Disease Control and Prevention's genetic reference material coordination program (*http://wwwn.cdc.gov/clia/Resources/GetRM/default.aspx*, last accessed September 9, 2016) and the Genome in a Bottle Consortium cell line repository (eg, the National Institute of Standards and Technology pilot genome reference material 8398, *http://jimb.stanford.edu/giab-news/2016/5/31/the-pilot-giabnist-reference-material-8398-is-now-available-1?rq=pilot*, last accessed September 9, 2016)[14] provide particularly useful resources. Despite their advantages, cell lines have limitations for MBPT. First, given the cost and time required, it is unreasonable to envision the production of a cell line for each of the wide range of somatically acquired variants found in cancer specimens (or, more broadly, the range of constitutional variants characteristic of inherited disorders). Second, cell lines that have been engineered to harbor specific variants often contain sequence artifacts related to the genetic manipulation,[15–17] and these artifacts can introduce non-physiologic complications into PT challenges (eg, unintended consequences for alignment and mapping). Third, mixtures of cell lines derived from different individuals introduce artifacts into the bioinformatics analysis of the sequence files as a result of non-physiologic patterns of single-nucleotide variants and allelic ratios (eg, if bioinformatics approaches are used to estimate specimen contamination).

Plasmid-based DNA constructs and chemically synthesized DNA fragments could likewise be used for MBPT. They can be designed to harbor a broad range of sequence variants, and can be used either alone or as so-called spike-in standards to introduce variants at a wide range of VAFs.[18,19]

REV 5.4.0 DTD ■ JMDI545_proof ■ 15 November 2016 ■ 5:43 pm ■ EO: JMD16_0140