APPROACHES TO THE VALIDATION OF HIGH THROUGHPUT SEQUENCING IN THE DIAGNOSTIC LABORATORY

Dr Ian Gardner

Canada Excellence Research Chair - Aquatic Epidemiology, Atlantic Veterinary College (3N), UPEI
550 University Ave., Charlottetown, Prince Edward Island, C1A 4P3, Canada
Tel: (+1-902) 620-5059; Fax: (+1-902) 620-5053; E-mail: iagardner@upei.ca

Genetic characterisation of infectious agents through high throughput sequencing (HTS) offers benefits including improved outbreak analysis (2); identification of novel strains to inform vaccine development, diagnostics and treatments; and greater insights into evolutionary microbiology.

Validation of the major technical components of HTS (sample preparation including enrichment steps, sequencing, and data analysis/interpretation), quality control procedures, proficiency testing of wet and dry lab steps, and reference material (DNA/RNA samples and data sets) are considered to be important components of overall quality management in a clinical setting (1). To my knowledge, there are no established guidelines for ensuring that HTS is fit-for-purpose and hence, reference laboratories and *ad hoc* working groups can play a key role in developing pathways for the validation of HTS-derived data that balance costs of validation, acknowledges that future technological developments will occur rapidly, and assumes increased completeness of informatics databases.

In the context of the OIE proposal, the most common application of HTS will likely be further characterisation of a pathogen detected in a primary assay. Assuming that HTS results are not used to establish a diagnosis, this application should be considered an adjunct test necessitating only analytical validation. A critical initial step is definition of the intended purpose of the HTS as the validation process must always be in the context of fitness for purpose. Estimation of analytical sensitivity of a primary isolate will be affected by extraction and purification methods for its nucleic acids whereas analytical sensitivity of unbiased sequencing will depend on the relative amounts of host and non-target microbial nucleic acid compared with target nucleic acid in the sample matrix. Analytical specificity may be approximated by the false positive rate which might reflect depth and uniformity of sequence coverage of the target region, base call quality scores etc., and the frequency of reagent contamination. Estimates of repeatability and reproducibility across sequencing platforms and computational software programs may need to be done on a limited basis because of cost considerations

Primary diagnostic (pathogen discovery) applications based on unbiased sequencing without prior knowledge of DNA content of samples are likely to increase in the future and hence, guidance is also needed on diagnostic validation. Parallel comparison of HTS results with those of pathogen-specific real-time PCR assays on the same set of representative field samples might be an appropriate choice.

1. GARGIS A.S., *et al*. (2012). Assuring the quality of next generation sequencing in clinical laboratory practice. *Nat. Biotechnol*., **30**, 1033–1036.

2. GILCHRIST C.A., *et al*. (2015). Whole genome sequencing in outbreak analysis. *Clin. Microbiol. Reviews*, **28** (3), 541–563.