Clinical Microbiology and Infection xxx (xxxx) xxx

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Commentary

Lessons learnt from the introduction of nanopore sequencing?

A. Heikema [1], W. de Koning [2], Y. Li [2], A. Stubbs [2], J.P. Hays [1, *]

[1)] Department of Medical Microbiology & Infectious Diseases, Erasmus University Medical Centre (Erasmus MC), Rotterdam, the Netherlands
[2)] Clinical Bioinformatics Unit, Department of Pathology, Erasmus University Medical Center (Erasmus MC), Rotterdam, the netherlands

ARTICLE INFO

Article history:
Received 18 February 2020
Received in revised form
26 March 2020
Accepted 23 May 2020
Available online xxx

Editor: G. Lina

In this commentary the authors comment briefly on their own experiences with the initial introduction of nanopore sequencing—Oxford Nanopore Technologies (ONT), Oxford, UK—for microbiota-based sequencing [1]. The commentary is based on joint experiences and collaboration between the Department of Medical Microbiology & Infectious Diseases and the Unit of Clinical Bioinformatics at the Erasmus University Medical Center Rotterdam (Erasmus MC). Initial pilot investigations into the implementation of nanopore sequencing began around May 2017.

The authors aim to provide feedback and suggestions to companies planning to launch future novel scientific technologies in this medical field. Furthermore, the authors do not focus on pure scientific aspects relating to nanopore sequencing (e.g. the original error rates generated by nanopore sequencing), but rather focus on more generally applicable topics.

This brief commentary is the authors1' own joint opinion, based on the use of nanopore for bacterial microbiota profiling, and does not necessarily reflect the opinions of other users of nanopore sequencing or of Oxford Nanopore Technologies itself. Finally, the authors appreciate that many advances have been made in nanopore sequencing technology since the technology was first brought to the market.

Only half of the picture?

One of the major claims associated with nanopore sequencing after its commercial introduction in 2015 was that scientists could perform sequencing in remote locations using a nanopore sequencing device ('extreme portability') Although true, the need for accessory devices - such as nucleic acid extraction devices and/or thermocyclers - may not have been fully emphasized in the marketing campaigns [2]. This is particularly relevant with respect to two aspects: (a) ancillary devices may not be as portable as the nanopore technology device itself (or if available may not be widely used), and (b) the quality of sequencing results obtained (this is true for all sequencing technologies) is dependent on the quality of the input DNA/RNA used; extreme portability may mean that devices are used in many different extreme conditions where the quality of the results obtained may not be easily reproducible or verifiable. Scientists should look rationally at the claims made by companies and understand the shortcomings of any new technology, including the need for additional ancillary devices and potential quality issues associated with, for example, sample processing [3].

Software upgrades and FAIR (findable - accessible - interoperable - reusable) data

In the rapidly evolving field of nanopore sequencing, regular software updates are being introduced by individuals and by ONT itself. However, the introduction of new software upgrades leads to problems for researchers (whichever new technology is being implemented). For example, improvements in bioinformatics pipelines may be made while a manuscript is in preparation or is under review, which may potentially impact on the final results and conclusions obtained. Although improvements/upgrades to software are always welcome, users of such software should be aware that their results may be impacted by future software improvements, including comparison with historically published articles. That said, the adoption and availability of FAIR data will help facilitate backwards comparison with historical articles, including nanopore-related research articles, as the historical data could be rerun using the most up-to-date software available. Another

* Corresponding author. J. P. Hays, Department of Medical Microbiology & Infectious Diseases, Erasmus University Medical Centre (Erasmus MC), Rotterdam, the Netherlands.
E-mail address: j.hays@erasmusmc.nl (J.P. Hays).

https://doi.org/10.1016/j.cmi.2020.05.035

potential problem associated with regular (although necessary) software updates is that FAST5 files, containing raw signal data, are required in order to perform base call sequencing. FAST5 files are large (we generated >250 Gb of FAST5 sequence data in a single Oxford Nanopore GRIDION run), and ~14x larger than base called FASTQ files- the file format normally saved by users for sequencing projects. As such, the storage of sequence runs containing large numbers of FAST5 files (as a safeguard against future software improvements for base calling) can be a challenge with respect to the amount of free data storage available and the cost of purchasing cloud storage or back-up hard drives on which to store the FAST5 data. This impacts on the quality of scientific results in a historical context, as opposed to impacting on scientific quality in a geographical context (as mentioned above).

Time to result

The generation of hundreds of thousands of reads and their subsequent base calling via, for example, Albacore and later Guppy with Oxford Nanopore Technologies' software may take considerable time (up to 48 h or more) depending on how deep the user wants the sequence data to be base called. This may be an experience associated only with microbiota sequencing, but the longer the user has to wait for results the less rapid nanopore sequencing becomes. It is therefore perhaps wise for companies with cloud services to regularly monitor the use of their cloud services so that they can take steps to proactively adapt capacity to the (increasing) demand of users as the popularity of their technologies increases. Further, although very rapid results may be obtained using nanopore sequencing, rapidity may not always be the driving factor for technology use; think for example about factors involving sequence coverage coupled to technological convenience (ease-of-use).

Output data format

Data generated with the easy-to-use bioinformatics software (see above) provided by ONT will frequently negate the need for further downstream processing via custom or other forms of available software. However, from our experience, nanopore-generated microbiota-based .csv output files require additional bioinformatics processing i.e., the creation of a specialized script, before the data can be analysed further. Specifically, the data in the downloaded .csv file contain TaxonIDs (number codes) instead of the actual taxonomic names (genus, species etc.) of the sequenced microorganisms. This means that after our sequencing experiments, users have to manually, or via specifically designed command line commands, 'decode' the TaxonIDs before further downstream processing can be performed. This step demands extra bioinformatics expertise, which reduces the rapidity and ease-of-use of the nanopore sequencing results. Companies should, therefore, consider that the format of output data should be convenient for end users without further need for custom command line programming, whilst maintaining accuracy by retaining the TaxonID with the genus and/or the species name.

Website design

Although the continued development of new kits and adaptations for novel (including nanopore sequencing) technologies are welcome, companies should ensure that their websites are easily navigable and should fully explain the range and intended use of the kits offered. Scientists need to be able to make intelligent choices about the potential use of new kit variants, with information presented in an easy-to-read (and searchable) format, i.e., what, where, when, how. In this respect, the website of ONT has much improved since its initial appearance. However, in the past, it was much more difficult to understand the specific intended application(s) of the various ligation, rapid, barcoding etc. kits that were available for purchase. Currently more of a standard practice, but as a reminder, it is useful for companies to hyperlink publication DOI (digital object identifier) references to their online kit descriptions so that potential users can identify the intended use of individual kits. It is also potentially useful to link kit descriptions to 'threads' within social media platforms. This process could be focused by linking threads to a specific, platform-based, community of users.

The ONT community

Nanopore sequencing is currently 'for research use only' and the potential development and feasibility of clinical diagnostic-based applications may be in the hands of a growing community of experienced nanopore sequencing users. Developing community-based channels on company websites is a good mechanism for developing new ideas and sharing scientific information promptly among the users of new technologies. The question, therefore, is how companies and users can best take advantage of such communities to facilitate the further development of new technologies towards regulatory approved clinical diagnostic use. For technologies that have a wide range of potential applications, perhaps one helpful solution is to focus their investment efforts on the 'critical mass' accumulated by different community 'threads' (including related scientific publications and social media presence) in order to help determine the most useful *Target Product Profiles* (TPPs – a planning tool that describes the desired characteristics or 'profile' of a product that is aimed at a particular target disease) [4].

Quality standards

One of the difficulties encountered when introducing new technologies into research and clinical environments is the inclusion of (universally) accepted standard quality control materials into the new research and diagnostic protocols being developed. Indeed, the issue of quality in sequencing (and currently especially microbiota-based sequencing) is a hot topic of concern, with at least one manufacturer (Zymo Research, USA) [5] currently offering free microbiota DNA and microorganism standards (January 2020) as part of a drive to promote universally accepted standardized microbiological materials for use as external and internal controls. The use of negative and positive controls per sequence device may seriously affect throughput and costs when device throughput is relatively low (e.g. 12 samples per flow cell). Of course, barcoding and mixing samples may help reduce this problem. However, the feasibility of this strategy depends on the sequencing depth required and the amount of time available to complete a sequencing run. Deciding on minimum quality standards and controls could be one of the main tasks delegated to a technology's online community (see also 'The ONT community' above).

Please cite this article as: Heikema A et al., Lessons learnt from the introduction of nanopore sequencing?, Clinical Microbiology and Infection, https://doi.org/10.1016/j.cmi.2020.05.035

Author contributions

JPH conceived the article. JPH and AH wrote the article. All authors took part in discussions relating to their experiences with nanopore sequencing, read the article and provided comments.

Transparency declaration

JPH reports grants from the European Union during the conduct of the study. AH, WK, YL and AS have no conflicts of interest to disclose.

References

[1] Oxford nanopore technologies. https://nanoporetech.com/.
[2] Krehenwinkel H, Pomerantz A, Prost S. Genetic biomonitoring and biodiversity assessment using portable sequencing technologies: current uses and future directions. Genes 2019;10.
[3] Fiedorova K, Radvansky M, Nemcova E, Grombiříková H, Bosák J, Černochová M, et al. The impact of DNA extraction methods on stool bacterial and fungal microbiota community recovery. Front Microbiol 2019;10:821.
[4] Target product profiles. https://www.who.int/research-observatory/analyses/tpp/en/.
[5] Zymobiomics microbial community standards. https://www.zymoresearch.com/collections/zymobiomics-microbial-community-standards.

Please cite this article as: Heikema A et al., Lessons learnt from the introduction of nanopore sequencing?, Clinical Microbiology and Infection, https://doi.org/10.1016/j.cmi.2020.05.035