Native-RNA-Sequencing throws its hat into the transcriptomics ring

Shobbir Hussain

Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom

Keywords: Nanopore sequencing; Oxford Nanopore Technologies, ONT, MinION, RNA-Seq, alternative splicing, epitranscriptome, epitranscriptomics

Correspondence: s.hussain@bath.ac.uk

Abstract

De novo sequence-level surveys of transcriptomes have previously relied on sequencing via a DNA-intermediate. While such methods can yield massive datasets, various problems mean that these do not always accurately reflect the true innate composition of transcriptomes. Enter Garalde et al. who present for the first time highly-parallel native RNA-Sequencing, with potentially disruptive future-implications for the transcriptomics field.

In order to understand comprehensively what aspects of a transcriptome are associated with and might regulate particular biological outcomes, it becomes necessary to perform a detailed and reliable survey of its composition. Understandably, investigators, wish this to be achieved within a single experiment, and at feasible monetary cost. Such needs were met, to some extent, now over two decades ago with the introduction of massively-parallel DNA sequencing techniques. Certain key characteristics of the data yielded by these methods however commonly frustrate transcriptomics researchers. For example, the very limited sequencing read-lengths obtained makes *de novo* profiling of alternate transcript isoforms susceptible to error, and largely dependent on biases of the computational algorithm chosen for read-reconstruction (1). Further, sequencing RNA via a DNA intermediate requires a number of enzymatic preparation steps (including but not limited to PCR-amplification) that can often introduce biases, meaning that the resulting data does not always accurately report the natural state of a transcriptome. More recently, newer highly-parallel sequencing methods such as that offered by Pacific Biosciences, has enabled RNA-Sequencing via fewer intervening preparation steps such as negating the need for PCR-amplification (2). Perhaps most importantly, single-molecule sequencing of full-length transcripts was also achieved on the platform, thus enabling unequivocal determination of splice-variant transcript isoforms (2). However, up until now, no method has enabled any degree of transcriptome-level sequencing by making measurements directly from native RNA strands.

The demonstration of using single protein nanopores for nucleotide sequencing of nucleic acid strands while in their native form was in fact made around two decades ago (3). The method currently works by embedding a protein pore with a nanoscale aperture into a synthetic membrane, and then applying a small potential difference across the pore-membrane assembly. The size of the resulting current will depend on the pore diameter, and will further be influenced by the degree via which any physical entity, such as a traversing nucleic acid strand, happens to block the aperture. If suitable motor enzymes are employed to ratchet the nucleic acid strand through the pore at a suitable speed, then characteristic signatures of current fluctuations as each base passes through the pore can be distinguished, thus enabling a reliable read of the sequence (4). Such principles have recently been scaled and further developed by Oxford Nanopore Technologies (ONT) who currently employ the bacterial CsgG amyloid secretion pore as the sensor. This has been used for highly-parallel native DNA sequencing (5), and has now been demonstrated by the developers for use in highly-parallel native RNA-Sequencing (6). For their described study, Garalde and colleagues at ONT used the 'MinION', which employs an array of 2048 protein nanopores divided between 512 channels, and enables recording of current/sequence in parallel from each channel. The study sequenced polyA+ RNA isolated from yeast, and reported >3 million native RNA reads pooled from 5

sequencing runs. Modal sequencing accuracy was ~90%, and ~79% of the reads (~2.7 million) successfully mapped to the *S.cerevisiae* transcriptome. Their analysis further showed that such reads had mapped accurately, and indicated that the majority of them spanned the full-length of annotated transcripts, thus revealing the method as a reliable analytical tool for transcriptomics-level investigations.

In addition to enabling single-molecule full transcript sequencing, further advantages of native RNA-Sequencing (which might be abbreviated to 'NatRNA-Seq') include that library preparation is exquisitely simple. This essentially consists of tethering an adaptor-motor complex to the polyA tail of RNA molecules, before applying these onto the sequencing array (Figure 1). A significant additional benefit of the method is its potential to directly sense covalent RNA base modifications. Indeed, high-resolution mapping of RNA modifications in complex transcriptomes, using high-throughput DNA sequencing methods, has in recent years led to considerable anticipation in the fledgling field of 'epitranscriptomics'. However, current progress may not be optimal, as DNA sequencing-based methods are not always reliable in accurately mapping the modifications, as exemplified by recent controversies in the field (7,8). Thus much more direct means of detecting modifications via NatRNA-Seq would likely be coveted by epitranscriptomics researchers. Encouragingly, previous work has already demonstrated the ability of nanopore-based sequencing to distinguish non-canonical RNA bases in synthetic molecules (9), and Garalde and colleagues indeed reinforce such findings in their study.

Despite the aforementioned benefits, there are significant challenges ahead, most notably with regards to current throughput of NatRNA-Seq: this needs to improve substantially for sufficiently comprehensive profiling of transcriptomic features. Throughput of ONT NatRNA-seq can currently be improved by performing a reverse transcription (RT) step prior to sequencing of RNA molecules, in order to disrupt their secondary-structure which can otherwise impede progression through nanopores. However, inclusion of RT does negate a few of the key advantages of NatRNA-seq over cDNA sequencing methods, and alternate solutions are thus worth considering. This may for example involve addition of some concentration of urea to the sequencing running buffer to disrupt RNA secondary-structures while maintaining pore-integrity (10). An additional potential way to improve throughput would be to identify and employ an enzyme which ratchets the RNA strands through nanopores at higher speeds; there is scope for this considering that the current motor for RNA strand-ratcheting operates at only ~1/6th the rate of that currently used for sequencing DNA molecules. The extent of potential throughput-improvement offered by the aforementioned measures is however quite limited, and significant advances will likely instead require more general developments in the technology. Currently an important issue with ONT-sequencing, is that the

protein nanopores employed deteriorate gradually during the sequencing process, and throughput therefore usually declines quite significantly during a run. Thus efforts to engineer a more stable sequencing nanopore is likely a significant component of ONT research and development endeavour. Ultimately, this may require a shift away from the use of proteins, to instead the employment of more stable synthetic solid-state nanopores, indeed currently an active area of investigation at ONT. Such advances would also most likely be associated with a substantial reduction in monetary cost for the end-user, thus enabling much wider adoption by 'omics' communities. Should these or analogous endeavours prove fruitful, we could witness a re-writing of the rule book concerning minimum expectations of the data-characteristics produced from many common types of transcriptome sequencing-based experiments.

Figure Legend

Figure 1. Simplified schematic of the ONT native RNA-Sequencing workflow. (A) Library preparation involves ligation of an adaptor pre-complexed with a motor enzyme to polyA+ RNA (for non-polyA+ species, the polyA tail may be enzymatically added). The ligated product is then applied to a flow cell containing the sequencing array. (B) The array consists of an assembly of hundreds to thousands of protein nanopores embedded into an electrically-inert synthetic membrane which is bathed in an electrolyte solution. A potential difference across the assembly is applied and the current amplitude across each channel is recorded by ONT software. During the process, a motor enzyme interacts with a nanopore and feeds its associated nucleic acid strand into it. (C) This will result in a deflection of the current recorded across the pore, with the characteristics of the continuing deflection depending largely on the precise biophysical and biochemical attributes of each nucleotide as they pass through. Trained computational software is used to convert the raw current signatures into nucleotide sequence, and this can be performed in real-time. Note that aspects of this process have been simplified for clarity and illustrative purposes; the interested reader should refer to Garalde et al. 2018 for full detail.

References

1. Steijger et al. 2013. Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods.* 2013 Dec; 10(12): 10.1038/nmeth.2714.

2. Sharon D, Tilgner H, Grubert F, and Snyder M. A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol.* 2013 Nov; 31(11): 1009–1014.

3. Akeson M., Branton,D., Kasianowicz,J.J., Brandin,E. and Deamer,D.W. (1999) Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.,* 77, 3227–3233.

4. Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat Biotechnol.* 2012 Feb 14;30(4):344-8. doi: 10.1038/nbt.2147

5. Carter J.-M., Hussain S. Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. *Wellcome Open Res.* 2017; 2:23.

6. Garalde, D., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods* doi: 10.1038/nmeth.4577.

7. Grozhik AV, and Jaffrey SR (2017). Epitranscriptomics: Shrinking maps of RNA modifications. *Nature* 551, 174–176 (09 November 2017). doi:10.1038/nature24156

8. Dominissini D., and Rechavi, G. (2017) Loud and Clear Epitranscriptomic m^1A Signals: Now in Single-Base Resolution. *Molecular Cell.* Volume 68, Issue 5, p825–826, 7 December 2017. DOI: http://dx.doi.org/10.1016/j.molcel.2017.11.029

9. Ayub M and Bayley H (2012). Individual RNA base recognition in immobilized oligonucleotides using a protein nanopore. *Nano Lett.* 2012 Nov 14;12(11):5637-43. doi: 10.1021/nl3027873.

10. Japrung, D., et al. (2010). Urea Facilitates the Translocation of Single-Stranded DNA and RNA Through the α-Hemolysin Nanopore. *Biophysics Journal.* Volume 98, Issue 3, Supplement 1, p44a, January 2010. DOI: http://dx.doi.org/10.1016/j.bpj.2009.12.255

Figure 1