

## TECHNOLOGY

## Expanding applications for nanopore sequencing

“ retains native modifications on the nucleic acid chains ”

Long-read sequencing platforms provide major opportunities relative to short-read sequencing. For example, the long reads facilitate the probing of repeat structures and the linking of distal features on the same chromosome or transcript. Additionally, the single-molecule nature without PCR-based amplification retains native modifications on the nucleic acid chains. Two new studies report advances for nanopore-based long-read sequencing: characterizing complex human transcriptomes, and enhanced analysis of human repetitive DNA regions.

In their study, Workman et al. optimized and performed nanopore sequencing of poly(A) RNA from the GM12878 human B lymphocyte cell line. This cell line already has a high-quality reference genome, which allowed the authors to assess the accuracy of the RNA sequencing runs. Although the per-read error rate is a primary hurdle in long-read sequencing, the authors achieved a per-base accuracy of ~86%, which is comparable to nanopore-based DNA sequencing and likewise can be increased through consensus analysis of multiple reads from the same region.

As one technical feature, the authors noticed frequent RNA truncations in the data. Rather than being merely a consequence

of reduced RNA stability relative to DNA, one major cause was aberrant ionic current spikes that terminated base calling prematurely. However, by identifying the aberrant signals, longer reads could be computationally reconstructed using sequence data upstream and downstream of the aberration.

The authors characterized numerous features of the transcriptome. They identified 33,984 transcript isoforms resulting from alternative splicing or transcription start sites of 10,793 genes. Importantly, the long reads enabled the combinations of exons to be determined throughout the transcripts, in contrast to the typical short-read identification of individual exon–exon junctions without long-range context. Similarly, the long reads facilitated analyses of allele-specific expression when transcript reads spanned multiple heterozygous sites, which allowed the alleles to be distinguished. Combining the analyses revealed some loci at which the two genomic alleles express different transcript isoforms. As a further example of long-range inferences, for *DDX5* transcripts the authors identified a correlation between poly(A) tail length and intron retention.

Finally, as proof of principle that nanopore sequencing can identify RNA modifications in complex transcriptomes, Workman et al. identified signatures of 6-methyladenosine (m<sup>6</sup>A) and A-to-I RNA editing.

Nanopore-based DNA sequencing can span long tracts of short tandem repeats (STRs); however, there is often uncertainty regarding the number of repeats detected. To improve accuracy, Giesselmann et al. designed CRISPR–Cas12a and CRISPR–Cas9 systems to make

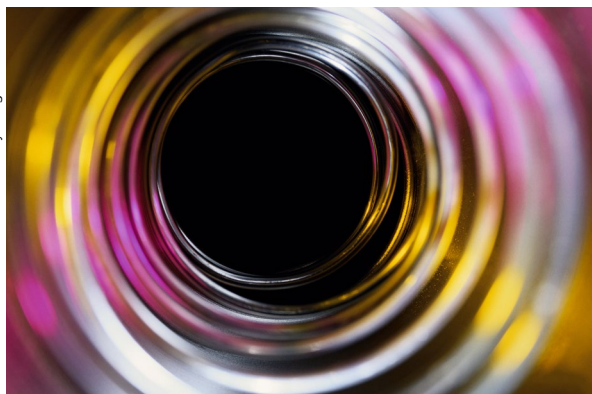
DNA cuts next to a repeat region of interest, which then facilitates ligation of sequencing adaptors for enrichment of these regions. Furthermore, they devised an algorithm called STRique that identifies the positions of the regions flanking STRs and then uses a hidden Markov model to count the number of STRs between them.

Giesselmann et al. applied their method to human cell lines in regions in which STR expansions are known to cause disease. They confirmed STR expansion in *C9ORF72* in a cell line from a patient with amyotrophic lateral sclerosis (ALS) and in *FMRI* in cell lines from a patient with fragile X syndrome. A model of pathogenesis of these diseases is that the expanded promoter STRs can lead to DNA-methylation-based silencing. Although DNA methylation analysis of the nanopore reads using a modification-adapted STRique algorithm confirmed the expected hypermethylation of *FMRI* STRs in cells from the patient with fragile X syndrome, no such hypermethylation was observed at *C9ORF72* STRs in cells from the patient with ALS, indicating that, at least in the cell line analysed, STR-methylation-based silencing of *C9ORF72* might not be the primary pathogenic mechanism.

These studies lay the groundwork for more widespread adoption of nanopore-based transcriptome and DNA repeat characterization, both of which can be combined with analyses of native modifications of nucleotides.

Darren J. Burgess

**ORIGINAL ARTICLES** Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019) | Giesselmann, P. et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. *Nat. Biotechnol.* **37**, 1478–1481 (2019)



Credit: Nick Koudis/PhotoDisc/Getty Images