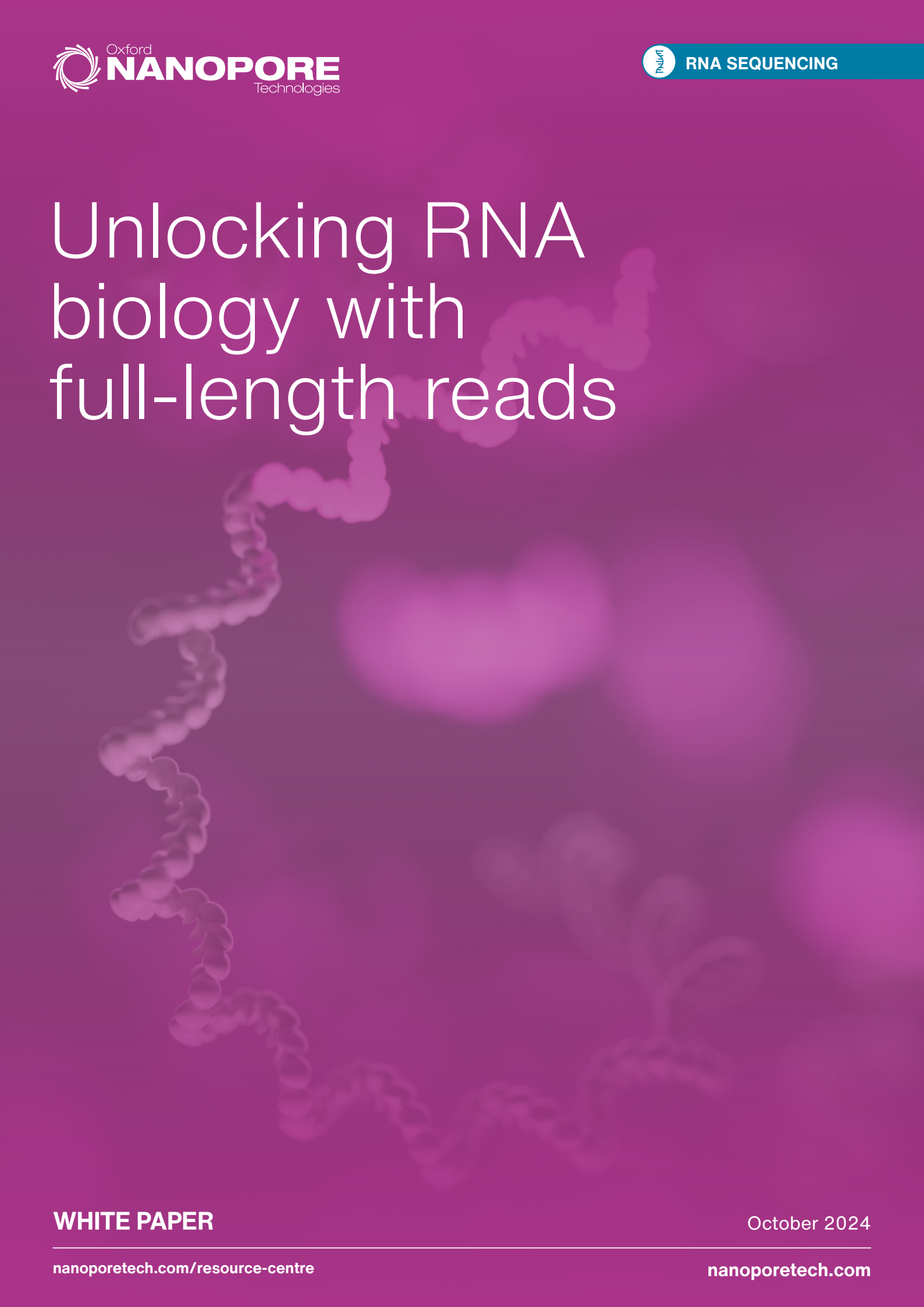
Oxford NANOPORE Technologies

RNA SEQUENCING

Unlocking RNA biology with full-length reads



WHITE PAPER

October 2024

nanoporetech.com/resource-centre

nanoporetech.com

Contents

Introduction

RNA sequencing has emerged as an indispensable tool in molecular biology, driving new insights into genomic function. This is particularly evident in the field of human and clinical research where transcript isoform expression and usage is a prominent source of variation between healthy and diseased tissues in many medical conditions[1,2]. In addition, RNA sequencing plays a pivotal role in the identification of fusion transcripts found in an expanding number of disorders[3–5].

> ***'Alternative splicing generates differing RNA isoforms that govern phenotypic complexity of eukaryotes. Its malfunction underlies many diseases, including cancer and cardiovascular diseases'***
>
> **Zhu, C. *et al.*[1]**

Sequencing of cDNA has also significantly aided viral pathogen characterisation and timely detection[6–8], dramatically improving time to result compared with the traditional technique of viral isolation in cell culture. Furthermore, unlike PCR-based virus detection methods, sequencing requires no prior knowledge of the viral genome, making it adaptable to rapid viral evolution[9].

The applications of RNA sequencing are not, however, restricted to the field of human and clinical research. The approach has also been utilised in agricultural settings, particularly the study of animal and plant disease, investigating plant resistance to various stressors, and molecular breeding[10,11].

Legacy sequencing technologies have unquestionably contributed significantly to scientific progress. However, these technologies possess inherent limitations that obscure vital information, hindering our complete understanding of the pivotal role of RNA. This white paper describes how nanopore sequencing overcomes these challenges to deliver the most comprehensive insights into RNA biology.

Solving the challenges of RNA sequencing

Obtaining and analysing full-length transcripts

Most eukaryotic transcripts are significantly longer than the 50–100 bp read lengths generated by legacy 'short-read' sequencing technologies. For example, human transcripts typically measure 1–2 kb in length, while the longest recognised processed human transcript, titin, extends beyond 100 kb[12]. The constraints on read length imposed by legacy RNA sequencing technologies present significant challenges in accurately assembling entire transcripts, especially where a read maps to more than one location (e.g. highly conserved stretches of sequence) (**Figure 1**). Reads that cannot be unambiguously aligned to a reference cannot be accurately quantified.

> ***'A series of problems [with short-read technologies], such as assembly difficulty and amplification biases, still make it a great challenge to accurately reveal the complexity of [the] transcriptome'***
>
> **Li, Y. *et al.*[13]**

With nanopore technology, full-length native RNA and cDNA molecules can be sequenced in single reads, effectively eliminating the challenge of multiple-locus alignment (**Figure 1**).

The long reads delivered by nanopore sequencing make the analysis of full-length transcripts easier than when using legacy sequencing platforms.

Revealing hidden biology through accurate isoform identification

The process of alternative splicing, whereby a precursor mRNA molecule can give rise to multiple mature mRNA molecules that differ in their exon content, 3'UTR, or polyadenylation sites, allows many proteins to be encoded by a single gene. In many cases, to unambiguously identify these unique mRNA isoforms, it is necessary to sequence several consecutive exons, which, due to their length, is beyond the capabilities of legacy short-read sequencing technologies (**Figure 2**). As a result, isoforms are reconstructed using computational techniques; however, a study by Steijger *et al.*[15] revealed that automated transcript assembly methods fail to identify all constituent exons in over half of the transcripts analysed. Furthermore, of those transcripts with all exons identified, over half were incorrectly assembled. These complications are further compounded where reads from highly similar transcripts, such as those of paralogous genes, are under investigation. Rare isoforms could remain altogether undetected[16].

> ***'[Long reads] make it possible to analyse transcripts and their isoforms without reconstructing them, or inferring their existence, from short reads; each sequence read simply represents its starting RNA molecule'***
>
> **Stark, R. *et al.*[14]**

The scale of this challenge for short-read technologies is stark. There are over 230,000 known human transcripts[17], with several eukaryotic genes known to encode hundreds to thousands of isoforms[18].

Long nanopore reads, comprising complete transcript isoform sequences, circumvent these issues with assembly, and their advantage in splice variant identification has been repeatedly demonstrated[18–20]. Of note, nanopore sequencing led to the identification of over 7,000 full-length isoforms of *Dscam1*, *'the most complicated alternatively spliced gene known in nature'*[18].

Reliable isoform resolution by nanopore technology makes *de novo* assembly of complex alternatively spliced transcriptomes possible, unlocking previously hidden genetic variation. For example, in a study of 88 human samples, Glinos *et al*. identified over 70,000 novel transcripts using long, full-length nanopore sequencing reads[21].

The benefits of nanopore sequencing also extend to the study of long non-coding RNAs (lncRNAs). This large class of RNA molecules, with lengths over 200 nucleotides, are involved in various regulatory and catalytic roles; however, due to the challenges of legacy sequencing technologies, their study has been largely overlooked. Researchers are now utilising nanopore sequencing reads to more comprehensively characterise and annotate lncRNA isoforms across multiple organisms, to help shed new light on their function and to explore their potential utility as biomarkers and drug targets in human disease[22–24].

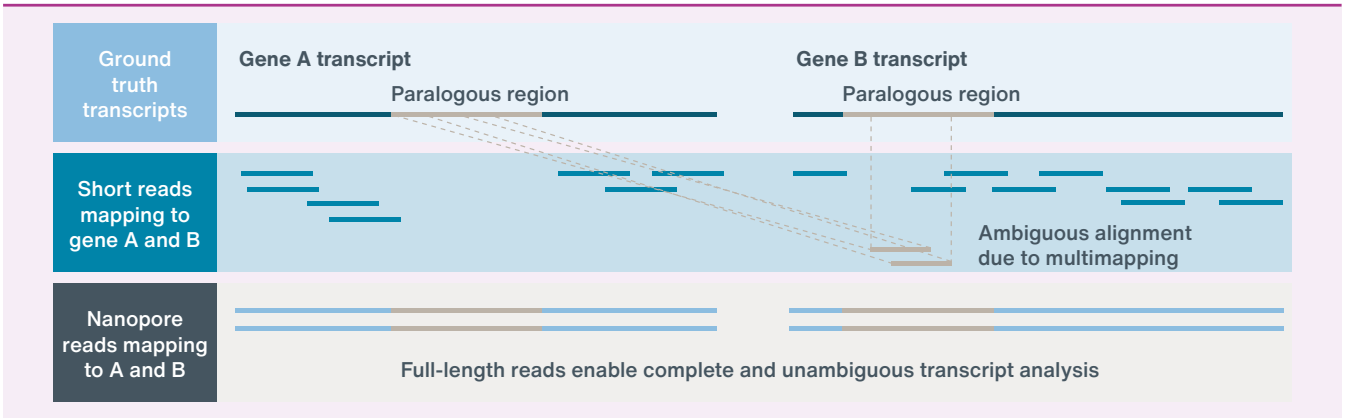


Figure 1
Transcript assembly comparison between legacy short-read sequencing technology and long nanopore sequencing reads. Nanopore sequencing reads can span full-length transcripts, effectively eliminating the challenge of transcript assembly and multimapping of repetitive or conserved paralogous regions faced by legacy sequencing technologies.

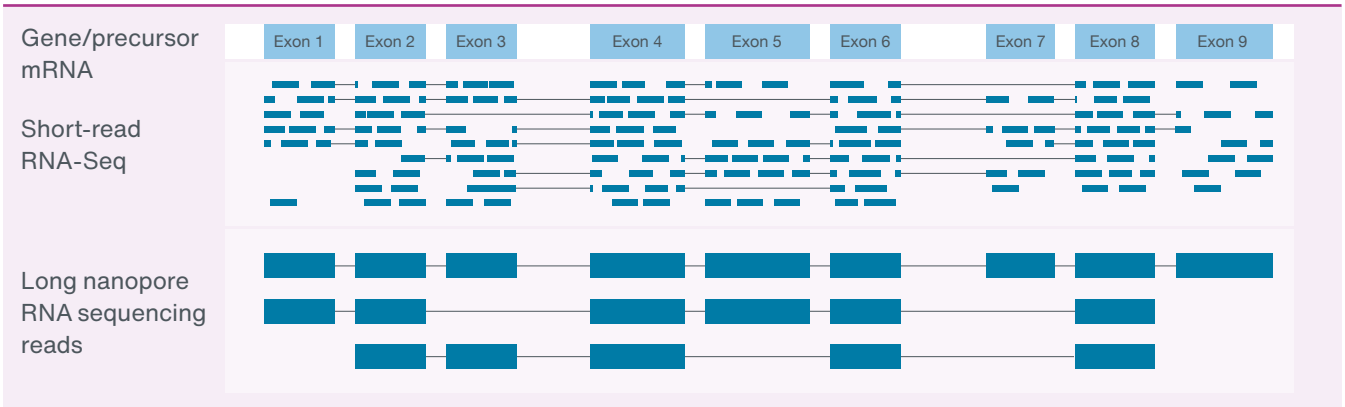


Figure 2
Alternative splicing can give rise to numerous mRNA isoforms per gene, which in turn can alter protein composition and function. The short reads generated by legacy RNA sequencing techniques lose positional information, making the correct assembly of alternative mRNA isoforms challenging. Long nanopore reads can span full-length transcripts, simplifying their identification.

Moving beyond basic gene expression

Due to the inherent read length limitations of legacy sequencing technologies, most transcriptomic studies focus on gene expression; however, this can miss important biology. For example, it is now known that individual transcript isoforms of the same gene can play entirely different, or even opposite, roles within a given cell[2]. According to Aguzzoli Heberle *et al.*, '*Knowing which tissues and cell types express each isoform is an important first step in understanding their function*'[2].

While some information on isoform usage can be inferred using short-read approaches, for reasons already described, they are fundamentally flawed for the generation of complete and accurate isoform expression data.

The facility of nanopore technology to sequence full-length transcripts in single reads overcomes these challenges to deliver more informative, isoform-specific insights than possible using legacy sequencing technologies.

Additionally, researchers have demonstrated that, due to their long length, far fewer nanopore sequencing reads are needed to deliver accurate gene expression data in comparison with short-read sequencing technologies[25,26]. In a study by Oikonomopoulos *et al.*, forty-fold fewer long nanopore sequencing reads, equating to approximately eight-fold fewer bases, were required to detect the same number of gene transcripts as provided by short-read sequencing technology.

> ***'Nanopore sequencing directly uncovered thousands of new RNA isoforms, many of which are expressed in a differentiation stage-specific manner'***
>
> **Ulicevic, J. *et al.*[27]**

In human clinical research, our understanding of the role of individual isoforms is critical to the potential development of more precise targets for disease treatment and diagnosis[2].

Discover more about isoform-level gene expression analysis: nanoporetech.com/gene-expression.

Single-cell sequencing at isoform resolution

While most transcriptome analyses to date have been performed on bulk cell populations, it is now possible to analyse the transcriptome at the single-cell level, unveiling previously hidden intricacies of cellular diversity and disease development[28–30]. However, because conventional single-cell assays combined with short-read sequencing typically capture only the 3' or 5' ends of transcripts, many single-cell studies have been limited to measuring expression at the gene level[31].

In contrast, nanopore sequencing enables researchers to measure both gene and isoform expression from a single sequencing run, maximising the biological data extracted from expensive single-cell isolations (**Figure 3**). Gene expression data is highly correlated between nanopore sequencing and short-read sequencing, while the nanopore data also uncovers cell-type-specific alternative splicing (**Figure 4**).

> ***'gene expression levels may only be revealing the tip of [the] iceberg of transcriptional diversities in cancer'***
>
> **Shiau, C.K. *et al.*[32]**

Although alternative sequencing platforms can deliver long reads, they currently cannot match the output, throughput, or ease of use of nanopore sequencing. As confirmed by Lebrigand *et al.*, '*Considering the vast amount of RNA molecules captured in current high throughput single-cell or spatial transcriptomics approaches, nanopore sequencing is a more attractive option to generate a sufficient amount of reads to reach the sequencing saturation needed for comprehensive transcript isoform and sequence heterogeneity exploration*'[33].

Furthermore, it is possible to prepare single-cell cDNA for nanopore sequencing in just three hours, delivering a significant timesaving over concatenation-based methods that currently require two days.

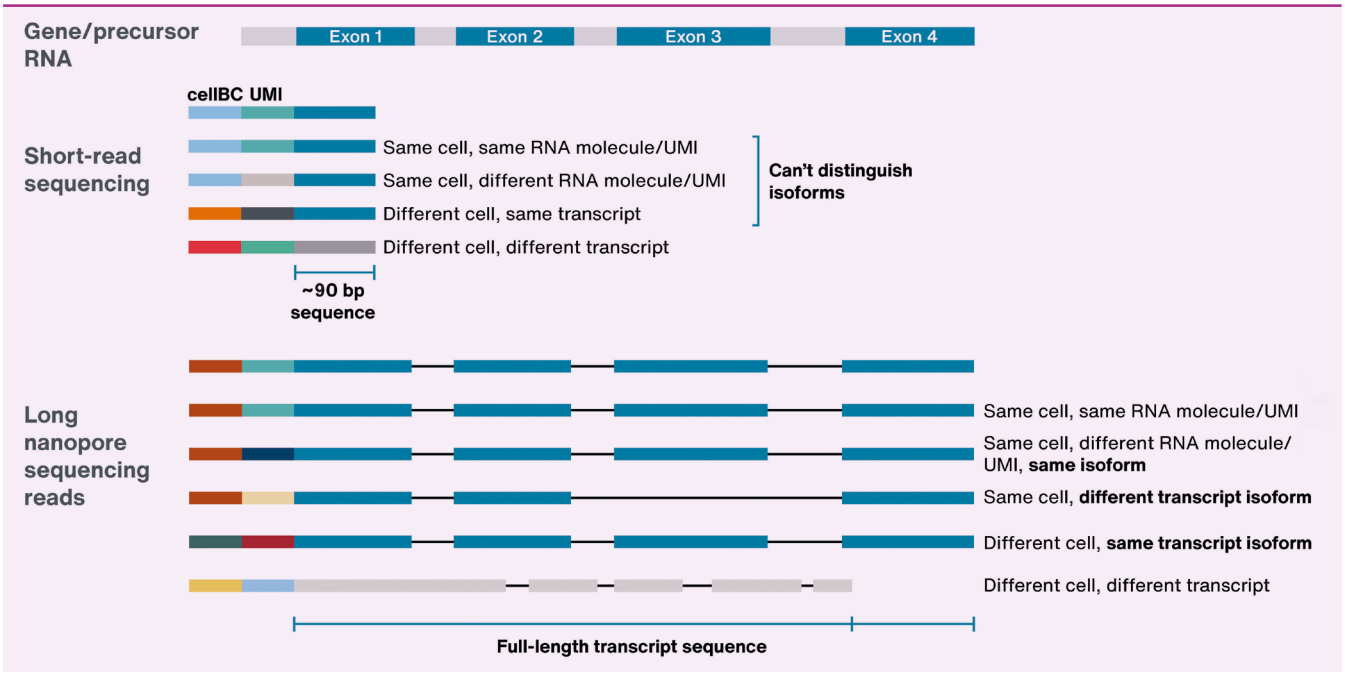


Figure 3

The power of single-cell sequencing comes from the facility to pool, sequence, and subsequently identify transcripts from hundreds of individual cells. All transcripts derived from the same cell can be identified using a unique cell barcode (cellBC). A unique molecular identifier (UMI), added to the original transcripts prior to amplification, further enables the identification of PCR duplicates that could impact transcript quantification, and can be used to generate more accurate consensus sequences. Short-read sequencing technologies only read approximately 90 bp of transcript sequence, precluding the identification of many transcript isoforms. In contrast, long nanopore sequencing reads can span complete transcripts, enabling comprehensive isoform-level gene expression analysis from single cells.

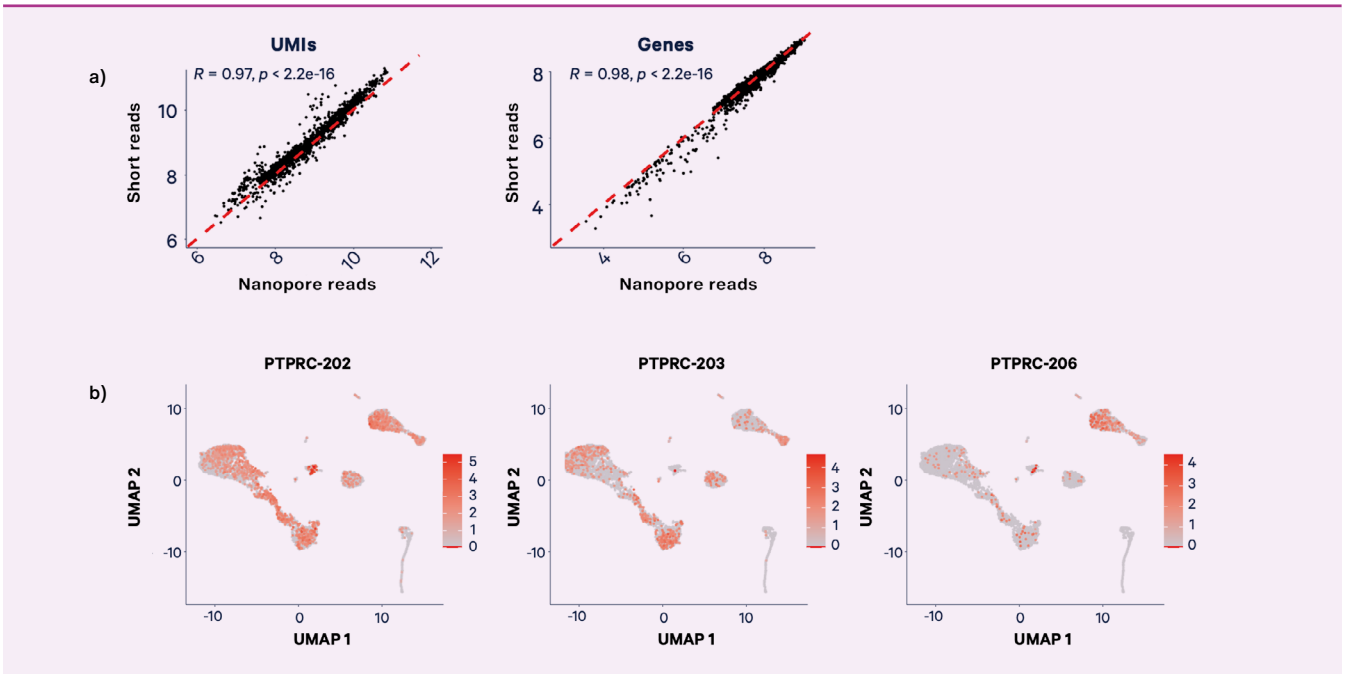


Figure 4

a) Nanopore single-cell sequencing reads show high correlation with legacy short-read sequencing data for UMIs and genes detected per barcode. b) Full-length nanopore sequencing reads further revealed cell-population specific expression of three alternative transcript isoforms of the *PTPRC* gene (which encodes CD45) that were not detectable using legacy short-read technology[34].

Discover more about nanopore single-cell sequencing: nanoporetech.com/single-cell.

Direct RNA sequencing — unbiased, full-length transcript and base modification analysis

Until recently, high-throughput, sequence-based analysis of RNA required the conversion of RNA to cDNA, which can introduce bias through reverse transcription and/or amplification. Research shows that PCR-amplified libraries tend to have reduced complexity when compared with the total mRNA pool[35]. Not all transcripts amplify with the same efficiency, causing drop-out of some RNA species and excessive amplification of others[35]. These issues can be exacerbated using legacy short-read sequencing technologies, which are known to exhibit GC bias, where sequences with low or high levels of GC content are underrepresented.

> ***'Sequencing native RNA has opened a whole new field of research'***
>
> **Smith, M.A.[36]**

An additional limitation of cDNA-based studies is that the process of PCR amplification loses all information on modified bases. Such base modifications are known to have a role in modulating the activity and stability of RNA and are therefore of increasing interest to researchers. Nanopore sequencing overcomes all these challenges through the facility for direct RNA sequencing — delivering unbiased, full-length, strand-specific native RNA sequences.

With nanopore sequencing, read length is only limited by the size of the RNA (or DNA) fragment presented to the pore, making it possible to sequence extremely long, full-length transcripts. The longest transcript processed by direct RNA sequencing currently stands at over 20 kb in length[37].

Direct RNA sequencing also enables the detection of epigenetic modifications alongside sequence data. It is now known that RNA methylation plays a key role in numerous disease mechanisms, including cancer[39], neurological disorders[40], and metabolic conditions[41], and the ability to routinely detect RNA modifications, alongside nucleotide sequence in a single, streamlined assay, promises to accelerate important discoveries in this area.

To date, the RNA modifications detected using nanopore sequencing technology include, but are not limited to, N6-methyladenosine (m6A)[35,42,43], pseudouridine (pseU)[23], inosine, and 5-methylcytosine (5mC)[35,44]. The ability to detect m6A is now integrated into the Oxford Nanopore sequencing software, meaning nucleotide sequences and modifications are called simultaneously with no additional downstream data processing required. See page 11 for more information on RNA modification detection.

> ***'Direct RNA sequencing enables the identification of m6A modification sites at the isoform level with single nucleotide resolution'***
>
> **Gleeson, J. *et al.*[40]**

In addition to research applications, direct RNA sequencing offers significant benefits to applied testing environments such as the development and quality control of biopharmaceuticals (e.g. mRNA-based drugs[45] and vaccines[46]), overcoming the inherent bias of legacy cDNA sequencing techniques and enabling simultaneous assessment of multiple characteristics (e.g. sequence identity, integrity, and purity) in a single assay. Find out more at: nanoporetech.com/biopharma.

Discover more about direct RNA sequencing: nanoporetech.com/rna.

Exploring the role of poly-A tails

A further benefit of nanopore sequencing — both cDNA and direct RNA — is the ability to accurately measure poly-A tail length. In eukaryotes, mRNA is augmented with a series of adenosine bases at the 3' end known as the poly-A tail. These tails can vary in size, with the longest being over 250 nucleotides in length and therefore beyond the typical analysis capabilities of short-read sequencing technologies[39,47]. Research suggests that poly-A tail length is an important factor in post-transcriptional regulation and further study may provide new insights into gene expression and disease[23,39].

> ***'The use of direct RNA nanopore sequencing highlights the ability to identify in one experiment the differential expression, modifications, polyA tail length, and unambiguous splicing isoform in the human transcriptome'***
>
> **Wang, D. *et al.*[23]**

Using nanopore sequencing, researchers have identified dynamic poly-A tail length distributions across the development stages of the model organism *Caenorhabditis elegans*[48]. Different poly-A tail lengths for alternative human transcript isoforms has also been described[49].

Rapidly identifying and characterising RNA viruses

Due to their lack of proofreading activity, RNA viruses mutate and evolve much faster than organisms with DNA-based genomes, allowing them to evade the host immune response and subsequent antiviral therapy. For this reason, most viral pandemics over recent years (e.g. COVID-19, Zika, Ebola, SARS, MERS) have been caused by RNA viruses. Offering streamlined workflows and real-time results, nanopore sequencing has been instrumental in developing effective genomic surveillance and public health strategies for emerging RNA viruses.

Common approaches to whole-genome characterisation of RNA viruses, include a tiled-amplicon methodology (suitable for DNA or RNA viruses and deployed widely for SARS-CoV-2) and direct RNA sequencing. Using direct RNA sequencing, which negates the need for reverse transcription and amplification, researchers at the Centers for Disease Control and Prevention, USA, were able to drastically reduce the time required for detection and characterisation of influenza virus from days to just a few hours[50]. In addition, the capability to directly detect RNA modifications is opening exciting new avenues of research[51].

Discover more about pathogen sequencing: nanoporetech.com/infectious-disease.

Cost-effective, scalable, and on-demand sequencing

Oxford Nanopore provides a range of devices that provide cost-effective, fully scalable, and on-demand sequencing to suit all research requirements (**Figure 5**).

The PromethION™ family of devices offer the highest sequencing outputs, with each flow cell typically delivering 100–200 Gb* of sequencing data. This makes them highly suited for isoform-level whole-transcriptome studies, including single-cell studies, where high-sequencing depth combined with long reads enables accurate identification of rare transcript isoforms. Running up to two or 24 high-output flow cells, respectively, the PromethION 2 and PromethION 24 devices provide the ultimate sequencing power and flexibility for isoform-level transcriptomics.

Nanopore sequencing devices allow each flow cell to operate independently, enabling multiple users to run different samples and experiments simultaneously. This flexibility further eliminates the need for sample batching, significantly accelerating result generation and allowing cost-effective analysis, regardless of project size or sample numbers. In contrast, legacy sequencing technologies require large samples numbers to achieve cost-efficiency, making them less suited for smaller-scale studies.

MinION™ Starter Packs (including a MinION device, flow cells, and sequencing reagents) provide all of the benefits of nanopore sequencing in a portable, low-cost format. MinION Flow Cells typically deliver up to 15–35 Gb* of data, ideal for isoform-level gene expression studies.

The benchtop GridION™ device can run up to five MinION Flow Cells independently or in parallel, offering a flexible, on-demand solution suitable and affordable for every lab.

Figure 5
Oxford Nanopore sequencing devices (from left to right): the portable MinION Mk1D; GridION, with capacity for five MinION Flow Cells; PromethION 2 Solo; PromethION 2 Integrated; and the high-throughput PromethION 24 device.

*Flow cell output will vary according to library type, run conditions, and other sample-specific factors. For more information on data output, visit: nanoporetech.com/products/specifications.

Discover more about Oxford Nanopore sequencing devices: nanoporetech.com/sequence.

Streamlined data analysis

Many tools and pipelines are available for the analysis of both cDNA and direct RNA nanopore sequencing data. Depending upon the aims of the study, sequencing reads are typically either mapped to the transcriptome or genome. Transcriptome mapping is more convenient for transcript identification and quantification but does not allow the discovery of novel isoforms. Alternatively, mapping to the genome allows more comprehensive analysis of splice variants and alternative isoforms; however, it is computationally more demanding.

With nanopore technology, sequencing reads are delivered in real-time and as standard output file formats, allowing analysis using a range of tools developed by the scientific community. In addition, Oxford Nanopore offers the fully supported EPI2ME™ analysis platform (**Figure 6**). EPI2ME provides researchers and bioinformaticians with easy access to best-practice analysis workflows, deployable locally or in the cloud and accessed via an intuitive desktop application or the command line.

A growing number of EPI2ME workflows are available, including:

- **wf-transcriptomes** — reference-guided or *de novo* transcript assembly and annotation from cDNA or direct RNA reads, plus differential gene expression and transcript usage analysis
- **wf-single-cell** — single-cell read mapping to a genomic reference, extraction of cell barcodes and UMIs, and generation of gene and transcript count matrices

In addition, as previously discussed, when using direct RNA sequencing, m6A modifications can be automatically called alongside the nucleotide sequence using MinKNOW™ — the software that controls nanopore sequencing devices. Furthermore, base modifications, including m6A, pseU, inosine, and 5mC can be detected using a standalone version of the Oxford Nanopore basecaller, which will be fully integrated into a future version of MinKNOW.

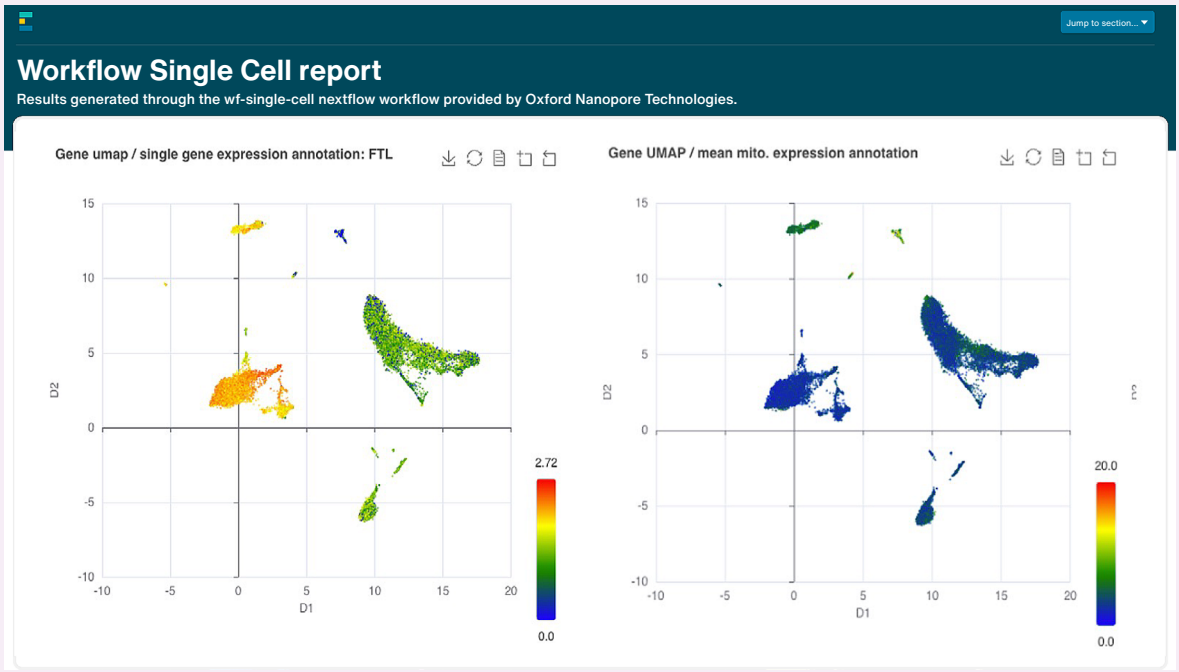


Figure 6
EPI2ME provides researchers and bioinformaticians with fully supported, best-practice data analysis workflows for a growing range of applications.

Discover more about analysing your nanopore sequencing data: nanoporetech.com/analyse.

CASE STUDY 1

Full-length RNA isoforms deliver new insights into human health and disease

Although it has long been understood that a single gene can generate multiple RNA isoforms that can result in different proteins, our knowledge of these isoforms — where they are expressed and how their functions may vary — remains limited. Legacy technologies such as short-read sequencing have prevented researchers from not only discovering and characterising all isoforms of a gene, but also quantifying expression and fully understanding their functions[52]. Using short-read sequencing, RNA must be fragmented, which leads to the problem of multimapping, impacting both identification of isoforms and differential expression analysis. Researchers have '*historically been forced to collapse all isoforms into a single gene expression measurement, [which is] a major oversimplification of the underlying biology*'[2]. Long nanopore sequencing reads can span full-length RNA transcripts, enabling accurate, unambiguous isoform identification and quantification.

In their recent publication, Aguzzoli Heberle *et al*. emphasised that long reads at high read depth are necessary to truly bridge the '*substantial gaps ... in our understanding of RNA isoform diversity*'[2]. With the aim of mapping medically relevant RNA isoforms in the human brain, the team used the cDNA-PCR Sequencing Kit and a PromethION device to perform whole-transcriptome sequencing of 12 post-mortem, aged, frontal cortex brain research samples: six with Alzheimer's disease (AD) and six cognitively unimpaired controls (CT), with a median of 35.5 million aligned reads per sample.

The team identified 7,042 genes expressing two or more RNA isoforms, 1,917 of which were determined to be medically relevant. Ninety-eight genes implicated in brain-related diseases were found to express multiple RNA isoforms, including AD genes such as *APP* (Aβ-precursor protein) with five isoforms, *MAPT* (tau protein) with four isoforms, and *BIN1* with eight isoforms. Several other genes implicated in other neurodegenerative diseases and neuropsychiatric disorders also expressed multiple RNA isoforms in the prefrontal cortex, including: *SOD1* (amyotrophic lateral sclerosis and frontotemporal dementia), *SNCA* (Parkinson's disease), *TARDBP* (involved in several neurodegenerative diseases), and *SHANK3* (autism spectrum disorder).

Using a strict threshold for high-confidence isoform identification, the team reported 428 new isoforms, 53 of which originated from medically relevant genes involved in brain-related diseases, including *MTHFS* (implicated in major depression, schizophrenia, and bipolar disorder), *CPLX2* (implicated in schizophrenia, epilepsy, and synaptic vesicle pathways), and *MAOB* (currently targeted for Parkinson's disease treatment).

> ***'Our study also uncovered 53 new RNA isoforms in medically relevant genes, including several where the new isoform was one of the most highly expressed for that gene'***
>
> **Aguzzoli Heberle, B. *et al.*[2]**

Five new spliced mitochondrial RNA (mtRNA) isoforms with two exons each were also identified. Explaining how surprising this was, the team revealed that all previously annotated human mitochondrial transcripts have only one exon; this has never been reported in human tissue before[2]. Highlighting that mitochondria are involved in many age-related diseases, the team shared how they are very interested in determining the function of these spliced mtRNA isoforms. Building on their new discoveries, they also identified RNA isoforms from genomic regions where transcription was not expected: 1,267 isoforms from 245 new gene bodies were reported. The median length was 1,529 nucleotides with 96.6% of isoforms only having two exons, which they suggest may be a feature of ageing in mammalian tissues.

The team shared that '*the most compelling value*' in using long nanopore sequencing reads is the ability to perform differential isoform expression analyses. Analysis of six AD and six CT samples revealed expression patterns associated with AD that were hidden when performing gene-level analysis. The team reported 176 differentially expressed genes and 105 differentially expressed RNA isoforms (**Figure 7**). Of the 105 isoforms, 99 came from genes that were not differentially expressed at the gene level. Using the gene *TNFSF12* as an example, the team showed that the *TNFSF12-219* isoform was significantly upregulated in AD research samples, whereas the *TNFSF12-203* isoform was significantly upregulated in control samples (**Figure 7**).

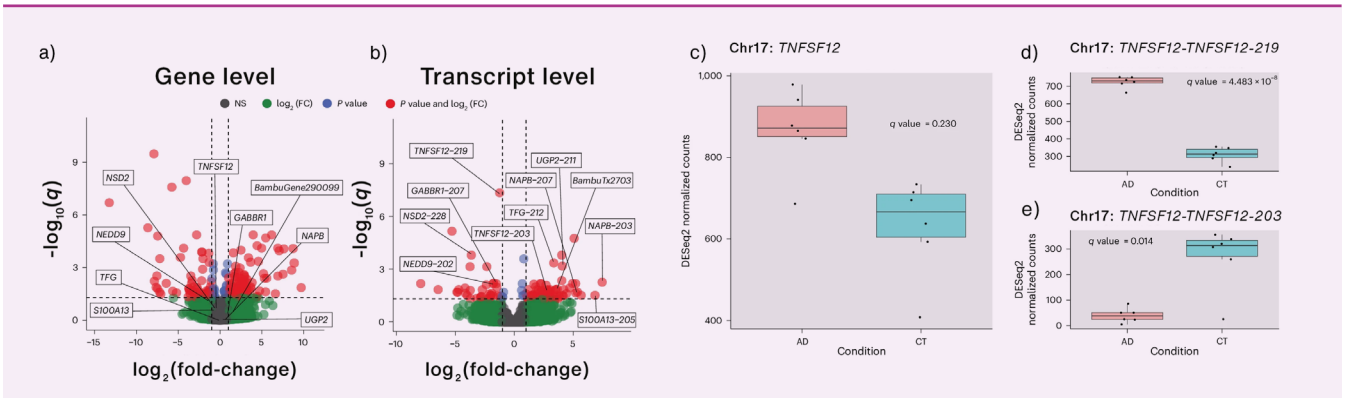


Figure 7

a) Differential gene expression and b) differential RNA isoform expression between AD research samples and CT samples. c) The *TNFSF12* gene was not significantly differentially expressed when collapsing all transcripts into a single gene measurement. d) *TNFSF12-219* was upregulated in AD research samples. e) *TNFSF12-203* was upregulated in CT samples. Figure taken from Aguzzoli Heberle *et al.*[2] and made available under Creative Commons License (creativecommons.org/licenses/by/4.0).

Understanding disease mechanisms and developing treatments require methods that offer '*substantial improvement over short-read sequencing*' approaches. Using long nanopore sequencing reads, Aguzzoli Heberle *et al*. demonstrated that a large proportion of medically relevant genes in the human frontal cortex expressed multiple RNA isoforms. Differential expression analysis of these isoforms can dive deeper, to reveal which isoforms are expressed in particular cell and tissue types, and potentially facilitate direct targeting of RNA isoforms for disease treatment.

CASE STUDY 2

Revealing hidden biology with isoform-level single-cell transcriptomics

To enable specialised functions, cells within an organism utilise the genome in different ways. Sequencing the transcriptome of each individual cell is an important step in revealing how this is achieved and is '*indispensable for understanding the underlying mechanisms of splicing and gene regulation*'[53]. Short-read sequencing has previously been used to characterise specific transcriptomic differences in single cells; however, the technology is limited in its ability to quantify RNA transcript isoforms because the transcripts must be fragmented, with just the 5' or 3' end being sequenced. In contrast, long nanopore sequencing reads can span complete transcripts, revealing full isoform diversity and enabling comprehensive isoform-level expression analysis in single cells.

Wang *et al.* from Baylor College of Medicine, USA, used nanopore sequencing to compile the first comprehensive characterisation of full-length transcript isoforms in individual mouse retinal cells[53]. The mouse retina is composed of over 130 unique cell types with their own distinctive transcriptomic profile, produced through alternative splicing of pre-mRNA. A comprehensive understanding of the RNA isoforms, splicing events, and the differential expression pattern at the single-cell level could be crucial in predicting the effect of genetic variants in retinal disorders.

> ***'One of the key advantages of long read [sequencing] is the improved ability of detecting transcript isoforms'***
>
> **Wang, M. *et al.*[53]**

Approximately 30,000 mouse retinal cells — two samples from wild-type retinas and two samples enriched in amacrine (AC) and bipolar cells (BC) — were profiled. The team reported high concordance (<98%) between short-read and nanopore datasets when comparing cell class assignments, although nanopore data '*identified an additional [bipolar cell] that was missed in the short-read data*'.

Using a PromethION device, 1.4 billion long nanopore reads were analysed, alongside 1.54 billion short reads. When sequenced at similar depths, both short-read and the nanopore datasets exhibited '*comparable sensitivity and high concordance in cell identification, clustering, and annotation*'. The team further observed that nanopore sequencing '*excelled in the precise identification of transcript isoforms*', with the median read length of approximately 1,000 nucleotides corresponding to the average size of full-length transcripts. The team identified 44,325 transcript isoforms, with approximately 40% being novel isoforms, which tended to be expressed at lower levels. It was suggested that this may be why these isoforms were undetected in previous studies and it was emphasised that single-cell nanopore sequencing '*greatly increased the number of isoforms detected*'.

Many genes were reported to display '*varying patterns of isoform usage among different cell classes and subclasses*'. Of the 44,325 transcript isoforms, 7,383 were pinpointed as cell-class specific, many of which were novel. Using long nanopore reads, the team discovered a '*common pattern*' where the major retinal cell classes expressed a '*combination of diverse isoforms rather than a single canonical isoform*', while '*intricate splicing variations*' between the two most abundant isoforms of a gene were frequently observed. However, similar to Aguzzoli Heberle *et al*. who also used long nanopore sequencing reads to identify diverse isoforms[2], Wang and colleagues found that it is not possible to identify cell-class-specific genes based solely on gene expression levels since the retinal cells expressed different isoforms even when their overall gene expression levels were not significantly different.

> ***'While transcript isoforms are often shared across various cell types, their relative abundance shows considerable cell-type-specific variation'***
>
> **Wang, M. *et al.*[53]**

Highlighting how the limitations of short-read sequencing has caused important biology to be missed, the team explained that using short-read technology it is '*necessary to rely on alterations in specific exons or splice junctions*', leading to '*suboptimal isoform reconstruction*'[53]. Single-cell nanopore sequencing comprehensively resolved the full-length transcriptome of the mouse retina, and in doing so uncovered some surprisingly complex gene fusion events. Within the mouse retinal cells, 1,055 intrachromosomal gene fusion transcripts were detected. Interestingly, although all fusions partners were on the same chromosome, they were not necessarily immediately adjacent to each other. Furthermore, some fusions exhibited alternative splicing. The researchers suggested that these findings highlight the complexity and flexibility of gene fusion events in the context of single-cell RNA sequencing data, and may present novel, clinically relevant insights.

> ***'the long-read sequencing approach demonstrated its reliability in detecting single-cell transcriptomes ... making it plausible to use ... exclusively for single cell RNA-seq in the future'***
>
> **Wang, M. *et al.*[53]**

The authors concluded that '*The integration of long-read sequencing with single-cell sequencing techniques holds the promise of filling the existing gaps in isoform information*'. They anticipate that their '*comprehensive atlas of full-length transcript isoforms*' mapped to individual mouse retinal cells will provide an '*invaluable resource for the community*'[53] and could help towards elucidating mechanisms of retinal disease.

CASE STUDY 3

Delivering '*unique insights*' into infection with direct RNA sequencing

Influenza is a highly contagious respiratory illness caused by influenza viruses. These relatively small RNA viruses (genome size: ~13.5 kb) are exceptionally adaptable and efficient at evading the host immune response, making producing a long-lasting vaccine against them challenging.

Over the years, efforts to study RNA viruses have been limited by the available techniques. Legacy sequencing technologies can only read short fragments of RNA, necessitating the use of often inexact computational assembly methods. In addition, the requirement for cDNA conversion not only introduces potential amplification bias but also overwrites base modifications, hiding potentially important epitranscriptomic insights. Now, using nanopore sequencing, native RNA can be directly sequenced, quantified, and analysed for base modifications, all in a single assay.

Wang *et al.*, from the University of Oklahoma, USA, used nanopore direct RNA sequencing to explore the changes in viral and host transcripts, and the host immune response, during influenza infection[23]. Using the Direct RNA Sequencing Kit and benchtop MinION sequencing device, the team analysed mRNA transcripts and non-coding RNA (both polyadenylated and non-polyadenylated) from influenza-exposed and mock-exposed human bronchial epithelial research samples. In a single assay, they were able to measure base modifications, transcript isoforms, poly-A tail length, and differential expression.

Differential transcript expression analysis suggested the influenza-exposed samples were at the early stages of infection, since the most significantly upregulated and downregulated transcripts related to immune response and virus entry.

This was further supported by the lack of viral RNA transcripts detected in the bronchial samples — consistent with the early stages of infection when the virus is bound to cells but has not yet entered or begun to replicate.

Direct detection of base modifications revealed 502 transcripts with m6A methylation in influenza-exposed samples, and 395 transcripts with m6A methylation in mock-exposed samples; however, the abundance of the majority of transcripts exhibiting differential m6A did not change. The team suggested that, although m6A methylation is an important regulatory mechanism of the transcriptional response to viral infection, it is unlikely to be the main mechanism regulating transcriptional changes in response to influenza in their model system.

The researchers further discovered that m6A methylation did not correlate with alternative splicing of mRNA in influenza-exposed samples. Unambiguous identification of complex, alternatively spliced isoforms is possible using nanopore sequencing as entire strands of native RNA are read in a single continuous read.

Of note, the team identified a significant increase in expression of one of the eight transcript isoforms of stimulated gene 12 (*ISG12*). This gene is known to suppress viral infection and exhibits complex alternative splicing. The researchers highlighted that

'*such complex alternative splicing is difficult to identify in short-read sequencing with data from reads 100–200 bp in length*'.

> ***'nanopore data unambiguously identifies the isoforms of transcripts that have multiple alternative splice sites'***
>
> **Wang, D. *et al.*[23]**

Methylation did not correlate with poly-A tail length either. An abundance of genes (410) associated with immunity and immune regulation showed a change in poly-A tail length. However, those with the greatest change in gene expression were less likely to show a change in poly-A tail length. This led the team to hypothesise that there are distinct, independent mechanisms of transcriptional and post-transcriptional regulation of host response genes during influenza infection.

Interestingly, the team observed increased methylation in response to influenza exposure in two long non-coding RNAs (lncRNAs) — CHASERR and LEADR, which were recently discovered to have roles in the immune response. Explaining that increased methylation could change the shape of the lncRNA, they predict methylation may disrupt the molecular interactions through which these lncRNA perform regulatory functions. In addition, the expression levels of several small nucleolar RNAs (snoRNAs), a type of non-coding RNA, were decreased in response to influenza.

The snoRNAs guide pseudouridylation. Nanopore sequencing revealed a corresponding decrease in pseudouridylation in two novel lncRNAs.

Although further research is necessary to fully understand the mechanisms through which noncoding RNA and RNA modifications fine tune the host response to influenza infections, the team highlighted how direct RNA sequencing using nanopore technology revealed '*unique insights*'.

> ***'The ability to identify multiple modifications on a long read of an individual RNA molecule surpasses other techniques'***
>
> **Wang, D. *et al.*[23]**

Furthermore, Hewel *et al*.[43] evaluated the latest nanopore RNA chemistry and basecalling model integrated into MinKNOW, the software onboard nanopore sequencing devices. Using a PromethION device, they sequenced standardised RNA test samples from HEK293T and Universal Human Reference RNA (UHRR) cell lines, and human blood research samples. Reporting increased sequence accuracy and throughput, the team also highlighted the ability to detect native RNA modifications '*in a close to real time*' (**Figure 8**). The researchers shared how the new chemistry opens up '*very intriguing possibilities*', including the potential for RNA diagnostics, and possible future clinical applications observing RNA modification disorders as a routine measure.

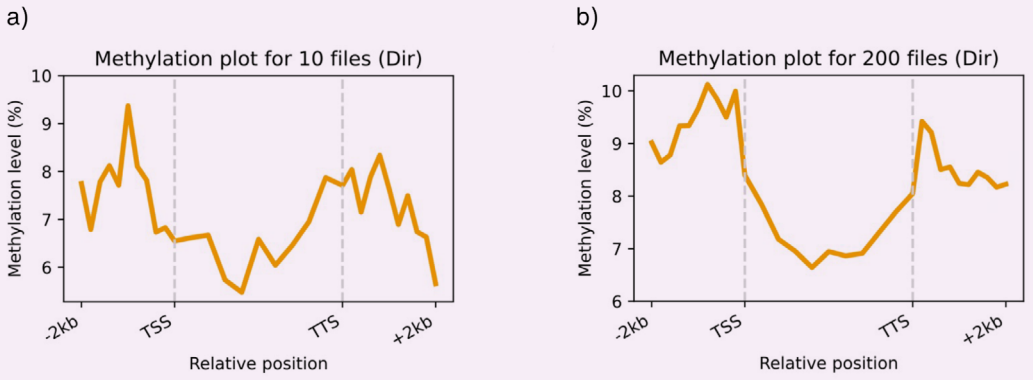


Figure 8
Direct (dir) detection of m6A in blood research samples for sequences up to 2 kb upstream of the transcription start site (TSS), gene bodies, and sequences up to 2 kb downstream of the transcription termination site (TTS). Plots were created in near real-time after every a) 10 files and b) 200 files output by MinKNOW. Figure taken from Hewel *et al.*[43] and made available under Creative Commons License (creativecommons.org/licenses/by-nd/4.0).

Summary

Although legacy short-read sequencing methods have advanced our understanding of many biological processes, they come with notable limitations, such as sequence bias and ambiguity in both the assembly of full-length transcripts and isoform quantification.

Nanopore technology provides full-length transcript isoforms as standard, enhancing the power of transcriptomic studies to explore the abundance and function of individual transcripts. Portability, rapid library preparation protocols, and real-time analysis pipelines offer further advantages, particularly in the area of viral pathogen surveillance[8].

'Individual RNA isoform expression has been overlooked due to technical limitations of short-read sequencing. Now with long-reads we can more accurately quantify the complete set of RNA isoforms'

Aguzzoli Heberle, B. *et al.*[54]

The advent of direct RNA sequencing offers a new avenue of transcriptome investigation, eliminating sequence bias observed with legacy technologies and enabling accurate identification of RNA modifications[35], without additional sample preparation steps or sequencing runs. These features, combined with the scalability offered by nanopore sequencing devices, make the technology, to quote Dr Marinov from Indiana University Bloomington, '*the most exciting development in transcriptomic research*'[55].

About Oxford Nanopore Technologies

Oxford Nanopore's goal is to enable the analysis of anything, by anyone, anywhere. The company offers the only sequencing technology to combine scalability — from portable to ultra-high-throughput formats — with real-time data delivery and the ability to elucidate accurate, rich biological data through the analysis of short to ultra-long fragments of native DNA or RNA.

Nanopore technology delivers full-length transcripts along with the ability to analyse poly-A tails and base modifications, offering the most complete single-platform solution for transcriptomics studies.

Affordable, scalable devices with independent flow cells eliminate the need for sample batching — providing every lab and researcher, regardless of project size, with immediate and cost-effective access to the most comprehensive insights into RNA biology.

Find out more about the value of full-length RNA sequencing at: www.nanoporetech.com/rna.

References

1. Zhu, C. et al. Single-molecule, full-length transcript isoform sequencing reveals disease-associated RNA isoforms in cardiomyocytes. *Nat. Commun.* 12(1): 4203 (2021). DOI: https://doi.org/10.1038/s41467-021-24484-z

2. Aguzzoli Heberle, B. et al. Mapping medically relevant RNA isoform diversity in the aged human frontal cortex with deep long-read RNA-seq. *Nat. Biotechnol.* 10.1038/s41587-024-02245-9 (2024). DOI: https://doi.org/10.1038/s41587-024-02245-9

3. Davidson, N.M. et al. JAFFAL: detecting fusion genes with long-read transcriptome sequencing. *Genome Biol.* 23: 10 (2022). DOI: https://doi.org/10.1186/s13059-021-02588-5

4. Kerbs, P. et al. Fusion gene detection by RNA-sequencing complements diagnostics of acute myeloid leukemia and identifies recurring *NRIP1-MIR99AHG* rearrangements. *Haematologica* 107(1): 100–111 (2022). DOI https://doi.org/10.3324/haematol.2021.278436

5. Felton, C., Tang. A.D., Knisbacher. B.A., Wu, C.J., and Brooks, A.N. Detection of alternative isoforms of gene fusions from long-read RNA-seq with FLAIR-fusion. *bioRxiv* 502364 (2022). DOI: https://doi.org/10.1101/2022.08.01.502364

6. Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530: 228–232 (2016). DOI: https://doi.org/10.1038/nature16996

7. Tyson, J. R. et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* 283077 (2022). DOI: https://doi.org/10.1101/2020.09.04.283077

8. Rambo-Martin, B.L. et al. Influenza A virus field surveillance at a swine-human interface. *mSphere* 5(1):e00822-19 (2020). DOI: https://doi.org/10.1128/msphere.00822-19

9. Young, K.T. et al. Randomly primed, strand-switching, MinION-based sequencing for the detection and characterization of cultured RNA viruses. *J. Vet. Diagn. Invest.* 33(2): 202–215 (2021). DOI: https://doi.org/10.1177/1040638720981019

10. Fracasso, A., Trindade, L.M., and Amaducci, S. Drought stress tolerance strategies revealed by RNA sequencing in two sorghum genotypes with contrasting WUE. *BMC Plant Biol.* 16(1):115 (2016). DOI: https://doi.org/10.1186/s12870-016-0800-x

11. He, S. et al. Nanopore direct RNA sequencing reveals the short-term salt stress response in maize roots. *Plants* 13(3):405 (2024). DOI: https://doi.org/10.3390/plants13030405

12. Bang, M.L. et al. The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circ. Res.* 89(11): 1065–1072 (2001). DOI: https://doi.org/10.1161/hh2301.100981

13. Li, Y. et al. (2023). Development of nanopore sequencing-based full-length transcriptome database toward functional genome annotation of the Pacific oyster, *Crassostrea gigas*. *Genomics* 115(5):110697. DOI: https://doi.org/10.1016/j.ygeno.2023.110697

14. Stark, R., Grzelak, M., and Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* 20(11): 631–656 (2019). DOI: https://doi.org/10.1038/s41576-019-0150-2

15. Steijger, T. et al. Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* 10(12):1177–84 (2013). DOI: https://doi.org/10.1038/nmeth.2714

16. Martin, J.A. and Wang, Z. Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12(10):671–82 (2011). DOI: https://doi.org/10.1038/nrg3068

17. Nurk, S. et al. The complete sequence of a human genome. *Science* 376(6588):44–53 (2022). DOI: https://doi.org/10.1126/science.abj6987

18. Bolisetty, M.T., Rajadinakaran, G., and Graveley, B.R. Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biol.* 16: 204 (2015). DOI: https://doi.org/10.1186/s13059-015-0777-z

19. Wu, H. et al. Nanopore long-read RNA sequencing reveals functional alternative splicing variants in human vascular smooth muscle cells. *Commun. Biol.* 6(1):1104 (2023). DOI: https://doi.org/10.1038/s42003-023-05481-y

20. Jones, E.F., Howton, T.C., Flanary, V.L., Clark, A.D., and Lasseigne, B.N. Long-read RNA sequencing identifies region- and sex-specific C57BL/6J mouse brain mRNA isoform expression and usage. *Mol. Brain* 17(1):40 (2024). DOI: https://doi.org/10.1186/s13041-024-01112-7

21. Glinos D.A. et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature* 608(7922):353-359 (2022). DOI: https://doi.org/10.1038/s41586-022-05035-y

22. Smith, M. Initial impressions of RNA004 and its application for functional profiling of cancer. Presentation. Available at: https://nanoporetech.com/blog/news-blog-my-research-60-seconds-improvements-direct-rna-sequencing-and-its-potential [Accessed: 16 August 2024]

23. Wang, D. et al. Nanopore direct RNA sequencing reveals virus-induced changes in the transcriptional landscape in human bronchial epithelial cells. *bioRxiv* 600852 (2024). DOI: https://doi.org/10.1101/2024.06.26.600852

24. Dyshlovoy, S.A. et al. Applications of nanopore sequencing in precision cancer medicine. *Int. J. Cancer* 10.1002/ijc.35100 (2024). DOI: https://doi.org/10.1002/ijc.35100

25. Bayega, A. et al. Nanopore long-read RNA-seq and absolute quantification delineate transcription dynamics in early embryo development of an insect pest. *Sci. Rep.* 11(1):7878 (2021). DOI: https://doi.org/10.1038/s41598-021-86753-7

26. Oikonomopoulos, S. et al. Methodologies for transcript profiling using long-read technologies. *Front. Genet.* 11, 606 (2020). DOI: https://doi.org/10.3389/fgene.2020.00606

27. Ulicevic, J. et al. Uncovering the dynamics and consequences of RNA isoform changes during neuronal differentiation. *Mol. Syst. Biol.* 20(7), 767–798 (2024). DOI: https://doi.org/10.1038/s44320-024-00039-4

28. Byrne, A. et al. Single-cell long-read targeted sequencing reveals transcriptional variation in ovarian cancer. *Nat. Commun.* 15(1):6916 (2024). DOI: https://doi.org/10.1038/s41467-024-51252-6

29. Cortés-López, M. et. al. (2023). Single-cell multi-omics defines the cell-type-specific impact of splicing aberrations in human hematopoietic clonal outgrowths. *Cell Stem Cell* 30(9): 1262–1281.e8. DOI: https://doi.org/10.1016/j.stem.2023.07.012

30. Thijssen, R. et al. Single-cell multiomics reveal the scale of multi-layered adaptations enabling CLL relapse during venetoclax therapy. *Blood* 140(20):2127-2141 (2022). DOI: https://doi.org/10.1182/blood.2022016040

31. Byrne, A., Cole, C., Volden, R., and Vollmers, C. Realizing the potential of full-length transcriptome sequencing. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 374(1786): 20190097 (2019). DOI: https://doi.org/10.1098/rstb.2019.0097

32. Shiau, C.K. et al. High throughput single cell long-read sequencing analyses of same-cell genotypes and phenotypes in human tumors. *Nat. Commun.* 14(1):4124 (2023). DOI: https://doi.org/10.1038/s41467-023-39813-7

33. Lebrigand, K. et al. The spatial landscape of gene expression isoforms in tissue sections. *Nucleic Acids Res.* 51(8):e47 (2023). DOI: https://doi.org/10.1093/nar/gkad169

34. 10x Genomics. Alternative transcript isoform detection with single cell and spatial resolution. Application note. Available at: https://nanoporetech.com/resource-centre/alternative-transcript-isoform-detection-with-single-cell-and-spatial-resolution [Accessed: 13 August 2024]

35. Garalde, D.R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15(3):201–206 (2018). DOI: https://doi.org/10.1038/nmeth.4577

36. Smith, M.A. https:/twitter.com/nanopore/status/1663213945383944193 (2023) [Accessed: 17 August 2024]

37. Timp, W. and Jain, M. Direct RNA cDNA sequencing of the human transcriptome. Presentation. Available at: https://nanoporetech.com/resource-centre/direct-rna-cdna-sequencing-human-transcriptome [Accessed: 17 August 2024]

38. Jain, M. Personal communication with Oxford Nanopore Technologies on 27 February 2018.

39. Kim, Y. et al. Nanopore-based native RNA sequencing of human transcriptomes reveals the complexity of mRNA modifications and crosstalk between RNA regulatory features. *bioRxiv* 603105 (2024). DOI: https://doi.org/10.1101/2024.07.11.603105

40. Gleeson, J. et al. Isoform-level profiling of m6A epitranscriptomic signatures in human brain. *bioRxiv* 578088 (2024). DOI: https://doi.org/10.1101/2024.01.31.578088

41. Rabolli, C.P. et al. Nanopore detection of METTL3-dependent m6A-modified mRNA reveals a new mechanism regulating cardiomyocyte mitochondrial metabolism. *Circulation* 149(16), 1319–1322 (2024). DOI: https://doi.org/10.1161/CIRCULATIONAHA.123.066473

42. Liu, H. et al. Accurate detection of m6A RNA modifications in native RNA sequences. *Nat. Commun.* 10(1):4079 (2019). DOI: https://doi.org/10.1038/s41467-019-11713-9

43. Hewel, C. et al. Direct RNA sequencing (RNA004) allows for improved transcriptome assessment and near real-time tracking of methylation for medical applications. *bioRxiv* 605188 (2024). DOI: https://doi.org/10.1101/2024.07.25.605188

44. Viehweger, A. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res.* 29(9):1545–1554 (2019). DOI: https://doi.org/10.1101/gr.247064.118

45. Zeglinski, K. et al. An optimised protocol for quality control of gene therapy vectors using nanopore direct RNA sequencing. *bioRxiv* 569756 (2023). DOI: https://doi.org/10.1101/2023.12.03.569756

46. Gunter, H.M. et al. mRNA vaccine quality analysis using RNA sequencing. *Nat. Commun.* 14, 5663 (2023). DOI: https://doi.org/10.1038/s41467-023-41354-y

47. Passmore, L.A. and Coller, J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat. Rev. Mol. Cell. Biol.* 23(2):93–106 (2022). DOI: https://doi.org/10.1038/s41580-021-00417-y

48. Roach, N. et al. (2020). The full-length transcriptome of *C. elegans* using direct RNA sequencing. *Genome Res.* 30(2): 299–312. DOI: https://doi.org/10.1101/gr.251314.119

49. Workman, R.E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* 16(12): 1297–1305 (2019). https://doi.org/10.1038/s41592-019-0617-2

50. Keller, M. Direct RNA sequencing of Influenza viral RNA using the MinION nanopore sequencer. Presentation. Available at: https://nanoporetech.com/resource-centre/direct-rna-sequencing-influenza-viral-rna-using-minion-nanopore-sequencer [Accessed: 17 August 2024]

51. Perlas, A. et al. Latest RNA and DNA nanopore sequencing allows for rapid avian influenza profiling. *bioRxiv* 582540 (2024). DOI: https://doi.org/10.1101/2024.02.28.582540

52. Page, M.L. et al. Surveying the landscape of RNA isoform diversity and expression across 9 GTEx tissues using long-read sequencing data. *bioRxiv* 579945 (2024). DOI: https://doi.org/10.1101/2024.02.13.579945

53. Wang, M. et al. Integrating short-read and long-read single-cell RNA sequencing for comprehensive transcriptome profiling in mouse retina. *bioRxiv* 581234 (2024). DOI: https://doi.org/10.1101/2024.02.20.581234

54. Aguzzoli Heberle, B. et al. Using deep long-read RNAseq in Alzheimer's disease brain to assess medical relevance of RNA isoform diversity. *bioRxiv* 552162 (2023). DOI: https://doi.org/10.1101/2023.08.06.552162

55. Marinov, G.K. On the design and prospects of direct RNA sequencing. *Brief. Funct. Genomics* 16(6):326-335 (2017). DOI: https://doi.org/10.1093/bfgp/elw043

Notes

Oxford Nanopore Technologies

phone +44 (0)845 034 7900

email support@nanoporetech.com

Oxford Nanopore Technologies

@nanopore

www.nanoporetech.com

WP_1047(EN)_V3_18Oct2024