frontiers in Molecular Biosciences

REVIEW
published: 02 August 2021
doi: 10.3389/fmolb.2021.711733

Isoform Age - Splice Isoform Profiling Using Long-Read Technologies

Ricardo De Paoli-Iseppi[†], *Josie Gleeson*[†] *and Michael B. Clark* *

Centre for Stem Cell Systems, Department of Anatomy and Physiology, The University of Melbourne, Parkville, VIC, Australia

Alternative splicing (AS) of RNA is a key mechanism that results in the expression of multiple transcript isoforms from single genes and leads to an increase in the complexity of both the transcriptome and proteome. Regulation of AS is critical for the correct functioning of many biological pathways, while disruption of AS can be directly pathogenic in diseases such as cancer or cause risk for complex disorders. Current short-read sequencing technologies achieve high read depth but are limited in their ability to resolve complex isoforms. In this review we examine how long-read sequencing (LRS) technologies can address this challenge by covering the entire RNA sequence in a single read and thereby distinguish isoform changes that could impact RNA regulation or protein function. Coupling LRS with technologies such as single cell sequencing, targeted sequencing and spatial transcriptomics is producing a rapidly expanding suite of technological approaches to profile alternative splicing at the isoform level with unprecedented detail. In addition, integrating LRS with genotype now allows the impact of genetic variation on isoform expression to be determined. Recent results demonstrate the potential of these techniques to elucidate the landscape of splicing, including in tissues such as the brain where AS is particularly prevalent. Finally, we also discuss how AS can impact protein function, potentially leading to novel therapeutic targets for a range of diseases.

Keywords: isoform, long-read sequencing, PacBio, Oxford Nanopore Technologies nanopore sequencing, single cell sequencing, alternative splicing, spatial transcriptomics, targeted RNA sequencing

OPEN ACCESS

Edited by:
Abdullah Kahraman, University Hospital Zürich, Switzerland

Reviewed by:
Alessio Colantoni, Italian Institute of Technology (IIT), Italy
Minna-Liisa Anko, Hudson Institute of Medical Research, Australia

***Correspondence:**
Michael B. Clark
michael.clark@unimelb.edu.au

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Protein and RNA Networks, a section of the journal Frontiers in Molecular Biosciences

Received: *19 May 2021*
Accepted: *19 July 2021*
Published: *02 August 2021*

Citation:
De Paoli-Iseppi R, Gleeson J and Clark MB (2021) Isoform Age - Splice Isoform Profiling Using Long-Read Technologies. Front. Mol. Biosci. 8:711733. doi: 10.3389/fmolb.2021.711733

INTRODUCTION

Alternative splicing (AS) enables the production of multiple RNA isoforms from single genes, greatly increasing both transcriptomic and proteomic diversity (Pan et al., 2008; Nilsen and Graveley, 2010). AS is the biological mechanism that controls which introns are removed from pre-mRNAs and which exons are joined to form the final messenger RNA (mRNA). Common AS events include skipped exons, retained introns and alternative 5′ and 3′ splice sites (Barbosa-Morais et al., 2012). It is now established that over 90% of multi-exon human genes undergo AS, and this prevalence highlights the biological importance of splicing events (Pan et al., 2008; Wang et al., 2008). Short-read sequencing approaches perform well at identifying AS of exons but are often unable to determine which full-length alternative isoforms are being expressed. The full extent of alternative isoform expression is only now starting to become clear with advances in long-read sequencing technologies that allow accurate determination of long-range exon connectivity (Sharon et al., 2013; Weirather et al., 2017).

The expression of different RNA isoforms can drive cellular differentiation, control cell functions and allow cells to respond to their environment (Roundtree and He, 2016). AS is highly regulated under normal conditions, while aberrant splicing contributes to various diseases including

neurological disorders, autoimmune disorders and the development of cancer (Emilsson et al., 2008; Lee and Young, 2013; Sui et al., 2014; Vitting-Seerup and Sandelin, 2017). Splice-altering variants that cause disease are more prevalent than previously anticipated (Soukarieh et al., 2016; Rhine et al., 2018), and it has been predicted that one third of all disease-causing variants lead to aberrant splicing (Lim et al., 2011). The increasing estimates of disorders attributable to aberrant splicing highlights the need for technologies that can accurately detect these changes at the isoform level.

In this review, we discuss current and emerging long-read sequencing methodologies for full-length RNA isoform detection and quantification including target enrichment, single cell and spatial transcriptomics approaches and ask how these can help us uncover the roles of alternative isoforms in health and disease.

SEQUENCING TECHNOLOGIES FOR DETECTING RNA ISOFORMS

Early methods for transcriptome-wide identification of expressed genes and isoforms involved cloning cDNA libraries into vectors followed by Sanger sequencing to ascertain isoform sequences. While laborious, such methods, combined with innovations such as cap-trapping and normalisation, identified hundreds of thousands of full-length isoforms in various cells and tissues, helping to reveal the complexity of the transcriptome (Carninci et al., 1996; Strausberg et al., 2002). Conversely, short-read second generation sequencing of cDNA offers high-throughput, fast and affordable measurement of gene expression levels but with trade-offs for the accurate identification and characterisation of isoforms (Trapnell et al., 2010; Engström et al., 2013; Steijger et al., 2013). The current rising popularity of long-read third generation sequencing methods relates in-part to their potential to combine the advantages of previous sequencing methods and profile full-length isoforms quickly and affordably.

Short-read sequencing (SRS) with Illumina is currently the most popular sequencing technology. SRS is a well-supported method for transcriptomics and is both high-throughput and affordable (Mortazavi et al., 2008). Sequencing instruments including the MiSeq (Illumina) or Ion Torrent (ThermoFisher Scientific) can produce reads of up to 600 nt long, however reads are most commonly within the 100–200 nt range. While SRS captures information such as splice sites and transcription start and end sites, short read methods struggle to determine how these features are combined into isoforms due to the fragmentation of RNA prior to sequencing. Therefore, while short reads accurately quantify gene expression, they often fail to identify the correct isoform from which the read originates as isoforms from the same gene are largely similar (Steijger et al., 2013; Soneson et al., 2016; Zhang et al., 2017). Long-read sequencing (LRS) technologies commercialised by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) have a distinct advantage over short-reads as they can reliably generate reads that cover the entire isoform. This removes the challenging task of reconstructing possible transcript isoforms from fragmented short reads and can improve our understanding of alternatively spliced isoforms of complex genes (Chen et al., 2019; Clark et al., 2020).

Single molecule, real-time (SMRT) sequencing developed by PacBio (California, United States) detects differently labelled dNTPs as they are incorporated into a DNA strand (Eid et al., 2009). The read length when using SMRT sequencing is primarily limited by the longevity of the DNA polymerase, with average read lengths of >20 kb now possible (Hon et al., 2020). To achieve high accuracy, DNA molecules are circularised prior to sequencing, allowing circular consensus sequencing (CCS), where the polymerase progresses around the circularised template multiple times. This allows highly accurate consensus sequences (<1% error rate), also called HiFi reads, to be generated from a portion of individual subreads (Wenger et al., 2019). The PacBio Iso-Seq method uses SMRT sequencing to generate a set of full-length transcript isoforms (Gonzalez-Garay, 2016). Iso-Seq can be performed on either unamplified or PCR amplified cDNA to detect and quantify isoforms (Ambardar et al., 2016), though some protocols include size fractionation of cDNA and separate sequencing of the fractions, increasing detection of longer isoforms but limiting opportunities for isoform quantification.

Nanopore sequencing, commercialised by ONT (Oxford, United Kingdom), uses an array of biological nanopores within a membrane that translocate nucleic acid under an electric current (Deamer et al., 2016; Jain et al., 2016). As nucleotides pass through the pore, they cause characteristic disruptions in the current that allow for their identification by basecalling software (Deamer et al., 2016). Recent improvements to the basecalling software that converts the raw current signal into nucleotide sequence have brought nanopore read accuracies into a similar range to SMRT sequencing (~90 to >99%) and both platforms continue to improve in this area (Wenger et al., 2019; Oxford Nanopore Technologies, 2021; Sahlin et al., 2021). Nanopore sequencing has no known length limit, with reads produced in excess of 2 Mb and length is primarily limited by the preparation and delivery of intact full-length sequences (Jain et al., 2016). Because sequencing is based on the detection of current changes caused by different nucleotides, nanopore sequencing can be performed on amplified and unamplified cDNA as well as native RNA (direct RNA sequencing). It can also detect epigenetic modifications, including RNA methylation, due to the characteristic current changes modified nucleotides create (Simpson et al., 2017; Lorenz et al., 2020).

Large numbers of reads are required to deeply profile the transcriptome and short-read methods will commonly generate 30–50 million reads per bulk sample. This is higher than the throughput currently obtainable with long-read platforms for the same experimental cost. However, PacBio can now generate up to 4 million HiFi reads on a Sequel II and ONT >50 million reads on a PromethION flow cell respectively. These continual increases in throughput have made comprehensive expression profiling feasible with long read platforms (Byrne et al., 2019).

Third generation, LRS technologies have been widely applied to discover and quantify gene isoforms across a range of species (including viruses, bacteria and plants), cell types and disease states. They have proved effective in characterising genes and

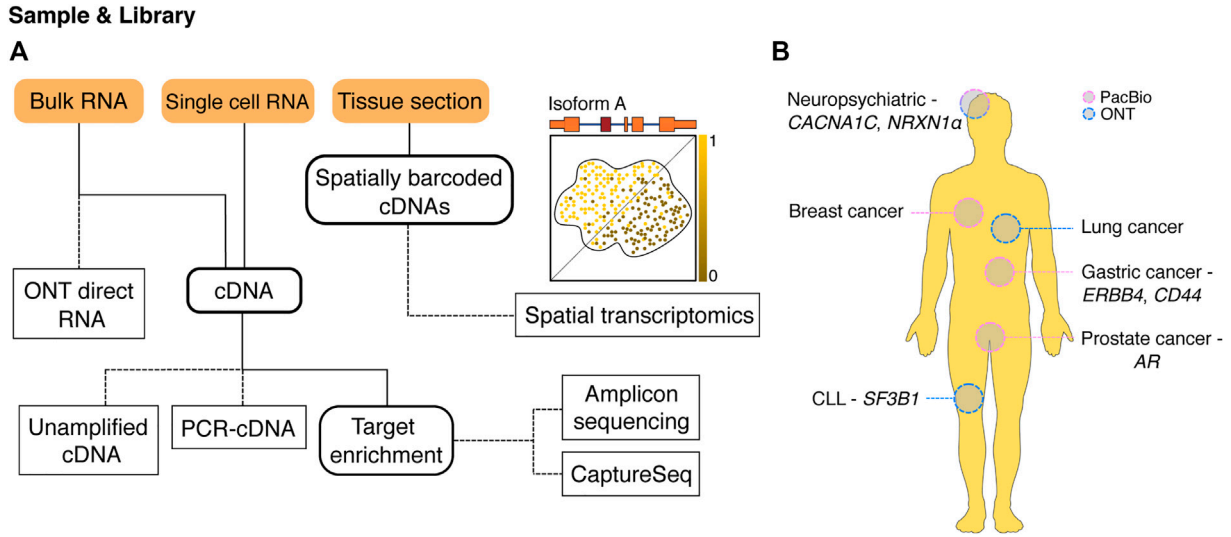

FIGURE 1 | Long-read methods for profiling isoforms. **(A)** An overview of available long-read methodologies for isoform characterisation. A variety of sample inputs (orange boxes) can be used with different sequencing methods (dashed edges) to answer a wide range of experimental questions. Intermediate experiment steps are shown with a bold outline. RNA from bulk samples can be either sequenced directly (ONT direct RNA) or reverse transcribed to cDNA, while RNA from single cells currently needs to be processed into cDNA before sequencing. Once present as cDNA, samples can be sequenced directly without PCR (unamplified cDNA) or after whole-transcriptome PCR amplification (PCR-cDNA) depending on starting amounts. cDNA can also serve as input for target enrichment techniques including PCR for specific gene isoforms of interest (amplicon sequencing) or capture of target cDNAs for sequencing (CaptureSeq). Long-read spatial transcriptomics starts with tissue sections placed on slides covered with spatially barcoded oligonucleotides. Cellular RNA is captured by nearby barcodes allowing generation of spatially barcoded cDNAs. Sequencing identifies spatial expression patterns of RNA isoforms within the tissue such as the hypothetical isoform A. **(B)** Selected examples where long-read sequencing methods have been used to examine isoforms and alternative splicing in a range of human diseases. AR, androgen receptor; CLL, chronic lymphocytic leukemia; PacBio, Pacific Biosciences; ONT, Oxford Nanopore Technologies.

isoforms in organisms with poorly annotated transcriptomes (Chen et al., 2017; Kim et al., 2019; Suryamohan et al., 2020), novel isoform discovery in well characterised organisms (Lagarde et al., 2017; Hardwick et al., 2019; Workman et al., 2019; Clark et al., 2020; Roach et al., 2020) and identifying changes in isoform profiles in disease (Asnani et al., 2020; Fujiyoshi et al., 2020; Tang et al., 2020). The portable nature of nanopore devices allows for use in the field and rapid response to emerging situations that can impact human health (Quick et al., 2015; Russell et al., 2018; Shaffer, 2019). Together these advantages are likely to see long-read methods continue to increase in popularity for expression and isoform profiling.

PROGRESS AND APPLICATIONS OF LONG-READ SEQUENCING FOR PROFILING SPLICED ISOFORMS

Long-read sequencing methods (**Figure 1A**) can cover entire transcripts within single reads, allowing for unambiguous identification of expressed gene isoforms. LRS can also be performed without any PCR or reverse transcription steps, reducing the amount of bias in expression quantification. These features allow for accurate quantification of both genes and isoforms, with results from LRS being comparable to or outperforming those from SRS (Oikonomopoulos et al., 2016; Garalde et al., 2018; Sessegolo et al., 2019; Soneson et al., 2019; Dong et al., 2020; Chen et al., 2021). Isoforms arising from genes with highly complex splicing patterns can be accurately detected with LRS, overcoming one of the major limitations of SRS (Clark et al., 2020; Chen et al., 2021).

In a recent benchmarking study, four sequencing methods (Illumina SRS, ONT: direct RNA, unamplified (direct) cDNA and PCR cDNA) were compared using five human cell lines (Chen et al., 2021). While SRS and LRS performed similarly for gene quantification, long-read methods outperformed Illumina for isoform level quantification. This was further highlighted in genes with a large number of similar isoforms where long-read quantification far exceeded short-read methods. Notably, there was a bias in the PCR cDNA data towards amplifying highly expressed genes, which was not observed in either the direct cDNA or direct RNA data. This suggested that although higher throughput can be achieved with amplification protocols, these may result in lower transcriptional diversity in the sequencing data. The authors found that thousands of genes were using multiple isoforms across the five cell lines. The most frequent difference between the major isoform and its alternatives was the use of alternative promoters, followed by exon skipping. The error rate of direct RNA sequencing is higher than in direct cDNA, however it was shown that isoform abundances were consistent across techniques, which indicated that the error rate is not a liability for isoform quantification. These results are consistent with those in a recent study using direct RNA sequencing that showed accurate isoform quantification and detection of differential isoform expression (DIE) in synthetic RNA spike-ins (Gleeson et al., 2020). Overall, the data presented

in Chen et al. (2021) suggests that long reads improve transcriptome profiling compared to short reads, while also allowing for additional analyses that are only made possible with long-read sequencing.

Various computational programs have been developed to produce a set of high-confidence isoforms from LRS data such as FLAIR, FLAMES, SQANTI and TALON (Tardaguila et al., 2018; Wyman et al., 2019; Dong et al., 2020; Tang et al., 2020). These programs correct splice junctions within reads and collapse identical transcripts into a set of unique high-confidence isoforms. This set of isoforms can then be aligned against current genome annotations for the discovery of novel isoforms. As a result, many studies have begun to reveal the extent to which human genome annotations are still incomplete (Soneson et al., 2019; Workman et al., 2019; Gleeson et al., 2020; Tang et al., 2020; Glinos et al., 2021). The percentage of novel unannotated isoforms found in these LRS studies is generally between 30 and 60%, further demonstrating how SRS misses many spliced isoforms. Although the false positive rate of novel isoform identification software remains unclear, studies on different cell lines have seen an overlap in the novel isoforms found (Glinos et al., 2021), and many novel isoforms have been validated with RT-PCR and Sanger sequencing (Tardaguila et al., 2018; Robinson et al., 2020; Uapinyoying et al., 2020). Additionally, these programs can utilise matched short-read splice junction information to inform splice junction correction and improve isoform quality (Tang et al., 2020). To further improve isoform detection and quantification, programs are now being developed specifically for estimation of isoform abundances from long-read data, such as NanoCount for direct RNA sequencing (Gleeson et al., 2020). Using knowledge of features unique to direct RNA, NanoCount improves quantification by more accurately identifying the isoform of origin for each read. Detection and quantification of isoforms with long reads has already been shown to outperform short reads, and this will only continue to improve as more tools are developed in this fast-paced field.

Long Read Profiling of Splicing in Disease

Alternative splicing contributes to diseases such as neurological disorders, autoimmune disorders and cancers (**Figure 1B**) (Emilsson et al., 2008; Lee and Young, 2013; Sui et al., 2014; Vitting-Seerup and Sandelin, 2017). Alternative splicing has also been shown to regulate the immune response to inflammation (Bhatt et al., 2012), and recent investigation of DIE using long-read sequencing has revealed a conserved mechanism of alternative first exon usage following inflammation (Robinson et al., 2020). In total, 50 novel isoforms with alternative first exon usage were identified from mouse macrophages, including a dominant novel isoform from the well-studied interferon inducible gene *Aim2* (Robinson et al., 2020). Understanding the role of AS in the inflammatory response will be important to improve our understanding of the molecular mechanisms that control inflammatory-regulated genes and the development of autoimmune disorders.

LRS has also been used to characterise viruses including Varicella-zoster (VZV), which causes chickenpox and shingles, within infected human neuronal cells (Braspenning et al., 2020). VZV establishes lifelong latency in neurons and its reactivation causes chronic pain later in life (Solomon et al., 2014). LRS revealed the architecture of the VZV transcriptome finding high levels of transcriptional complexity and alternative splicing (Braspenning et al., 2020). The ability to characterise viral transcriptomes with long-read sequencing has been especially useful throughout the COVID19 pandemic. Nanopore sequencing of the SARS-CoV-2 virus revealed the dynamic nature of transcription during its replication cycle (Chang et al., 2021). As well as identifying differential expression of subgenomic mRNA (sgRNA) transcripts during infection, novel sgRNAs containing non-canonical splice junctions were found that may have a role in enhancing viral protein production (Chang et al., 2021). These studies focusing on viral infection highlight how long-read sequencing enables the detection of full-length isoforms to reveal underlying viral mechanisms and further insight into the viral replication cycle within host human cells.

Alternative splicing and isoform switching are involved in the development of many cancers (Kahraman et al., 2020), and these mechanisms also contribute to cancer treatment resistance (Sciarrillo et al., 2020). For example, specific splicing patterns confer resistance to the cancer T-cell therapy CART-19 used to treat leukemia (Shah et al., 2019). Alternative splicing of CD19 transcripts was shown to play a central role in resistance to CART-19 immunotherapies, which led to further investigation of these transcripts using ONT direct RNA sequencing of human B-cells (Asnani et al., 2020). The study confirmed that an intron retention event caused premature termination of the CD19 transcript and the ablation of protein expression. Mutations in the splicing factor *SF3B1* are known to impact alternative splicing in several cancers (Furney et al., 2013; Tang et al., 2020). Using ONT cDNA sequencing and FLAIR, Tang et al. (2020) generated full-length isoforms from chronic lymphocytic leukemia (CLL) blood samples. This method identified 35 alternative 3′ splice-sites and global downregulation of intron retention events in *SF3B1* mutant CLL compared to wild-type (Tang et al., 2020).

A recent study of ten gastric cancer cell lines using PacBio's Iso-Seq technology identified approximately 39,000 novel isoforms, including for the known oncogenes *ERBB2* and *CD44*. Alternative promoters were frequently used in gastric cancer cells which often resulted in altered downstream CDS and 3′ UTRs (Huang et al., 2021). Full-length transcripts, using both ONT and PacBio reads, have also been sequenced to identify variants involved in treatment resistance in triple negative breast and lung cancer cell lines (Lian et al., 2019; Seki et al., 2019; Oka et al., 2021). Together, these studies highlight the need for long-read technologies to characterise cancer specific and/or cancer-causing isoforms with the potential to encode novel therapeutic targets or act as biomarkers (Kahles et al., 2018).

Long-Read Profiling of RNA Processing

The widespread interdependency of transcription initiation and subsequent mRNA splicing and processing was recently

established in human breast cancer cells using PacBio long-read sequencing, revealing that alternative transcription start sites had a significant impact on alternative splicing even across multiple exons and large distances (Anvar et al., 2018). These processes were tightly coupled in over 60% of genes with multiple transcripts and consistent across three other human cell types; brain, heart and liver. LRS is a promising method for investigating RNA regulation and processing, and these studies demonstrate our current incomplete understanding of coordinating mechanisms between transcription initiation and mRNA splicing.

Characterising the Role of Genetic Variation on Isoform Expression

A recent study advanced long-read transcriptional profiling into the realms of population-scale analysis on a large number of individuals (Glinos et al., 2021). LRS of 88 Genotype-Tissue Expression (GTEX) samples discovered almost 100,000 novel human isoforms, including support for over 4,000 novel isoforms previously identified by Workman et al. (2019). Over half of the highly expressed novel isoforms were only expressed in a single tissue, highlighting a potential role for these isoforms in tissue specific functions. An allele-specific analysis was used to study the effects of common and rare genetic variants on both RNA expression and splicing. Splicing quantitative trait loci were shown to mostly result in exon skipping, while expression quantitative trait loci modified not only expression levels, but also changes to the 5′ end of transcript structures (Glinos et al., 2021). The authors highlighted the importance of studying the transcriptome at the level of isoforms and splicing rather than at the gene level, which is now possible with long-read technologies.

TARGETED LONG-READ SEQUENCING

Long-read sequencing of whole transcriptomes has provided many insights into splicing and the role of RNA isoforms in health and disease. Expanding on these capabilities, a number of studies have now coupled LRS with technologies such as targeted sequencing of genes and isoforms of interest. Due to the wide dynamic range and often cell- or tissue-specific expression of RNAs, targeted methods are commonly needed for the detection and quantification of many isoforms (Mercer et al., 2014). This limitation to sequencing sensitivity affects LRS similarly to SRS and cannot easily be overcome with additional sequencing depth. RNA sequencing using CaptureSeq or PCR-amplicon sequencing has been successfully utilised to perform targeted long-read sequencing. These targeted approaches have been used to discover and annotate novel RNA isoforms, particularly of lowly expressed genes like long-noncoding RNAs (lncRNAs) (Lagarde et al., 2017; Hardwick et al., 2019); investigate splicing dynamics (Deveson et al., 2018); deeply profile disease gene isoform diversity and variation between tissues (Treutlein et al., 2014; Clark et al., 2020) and help confirm the impact of pathogenic variants on isoform expression (Helman et al., 2021). In this section we examine current targeted long-read sequencing methods and challenges in enriching for specific gene isoforms.

RNA Capture Sequencing Using Long Reads

Despite concerted efforts using SRS, many expressed genes and isoforms remain unannotated or incorrectly assembled, particularly for complex genes with many isoforms or lowly expressed genes such as lncRNAs. RNA CaptureSeq uses oligonucleotide probes to enrich for transcripts from genes or genomic loci of interest (**Figure 2A**). CaptureSeq can enrich one to 1,000's of genes allowing it to be applied to many gene sets of interest. Initially combined with SRS, it dramatically increased sequencing sensitivity, allowing for the discovery and quantification of new genes and isoforms (Mercer et al., 2014; Clark et al., 2015), however the identification of full-length isoforms was limited by the use of SRS.

Long-read RNA CaptureSeq (LCS) has now been successfully implemented in several studies to enable sensitive profiling of full-length isoforms. Lagarde et al. (2017) first developed LCS with PacBio sequencing to improve lncRNA annotations in human and mouse and thousands of novel full-length lncRNAs isoforms were discovered. The authors found that lncRNAs had at least twice as many isoforms as previously identified, and detection of lncRNA promoters and exonic structures was greatly improved (Lagarde et al., 2017). To expand the analysis of AS within lncRNAs, Deveson et al. (2018) used both short and long-read (PacBio) capture sequencing of human chr21, providing unprecedented sequencing depth to investigate transcription and splicing. Almost the entire non-repetitive length of chr21 was covered by spliced transcripts with very little exclusively intergenic space. Surprisingly, unlike coding transcripts which had defined alternatively spliced exons, almost every lncRNA exon could be alternatively spliced. The universal nature of AS for noncoding exons suggested lncRNA exons are modular in nature and that lncRNAs have a vast array of undiscovered alternative isoforms (Deveson et al., 2018).

Many genomic regions have been implicated in neurological disorders by genome-wide association studies (GWAS). However, many of these genomic regions are annotated as intergenic (Buniello et al., 2019). Short-read CaptureSeq has identified widespread noncoding transcription in these "intergenic" GWAS loci (Bartonicek et al., 2017). To further investigate this, long-read CaptureSeq with ONT was used to interrogate genomic regions implicated in neuropsychiatric disorders (Hardwick et al., 2019). LCS identified 109 novel high-confidence multi-exonic transcripts from these regions, along with new alternative isoforms of known brain genes, which were predicted to encode for novel protein variants (Hardwick et al., 2019). Gene and isoform discovery in "intergenic" GWAS regions will help facilitate functional genomic studies of how these regions contribute to disease.

A key limitation of CaptureSeq is the high cost of the oligonucleotide capture probes. Recently, Sheynkman et al. (2020) developed ORF CaptureSeq (OCS), a method to

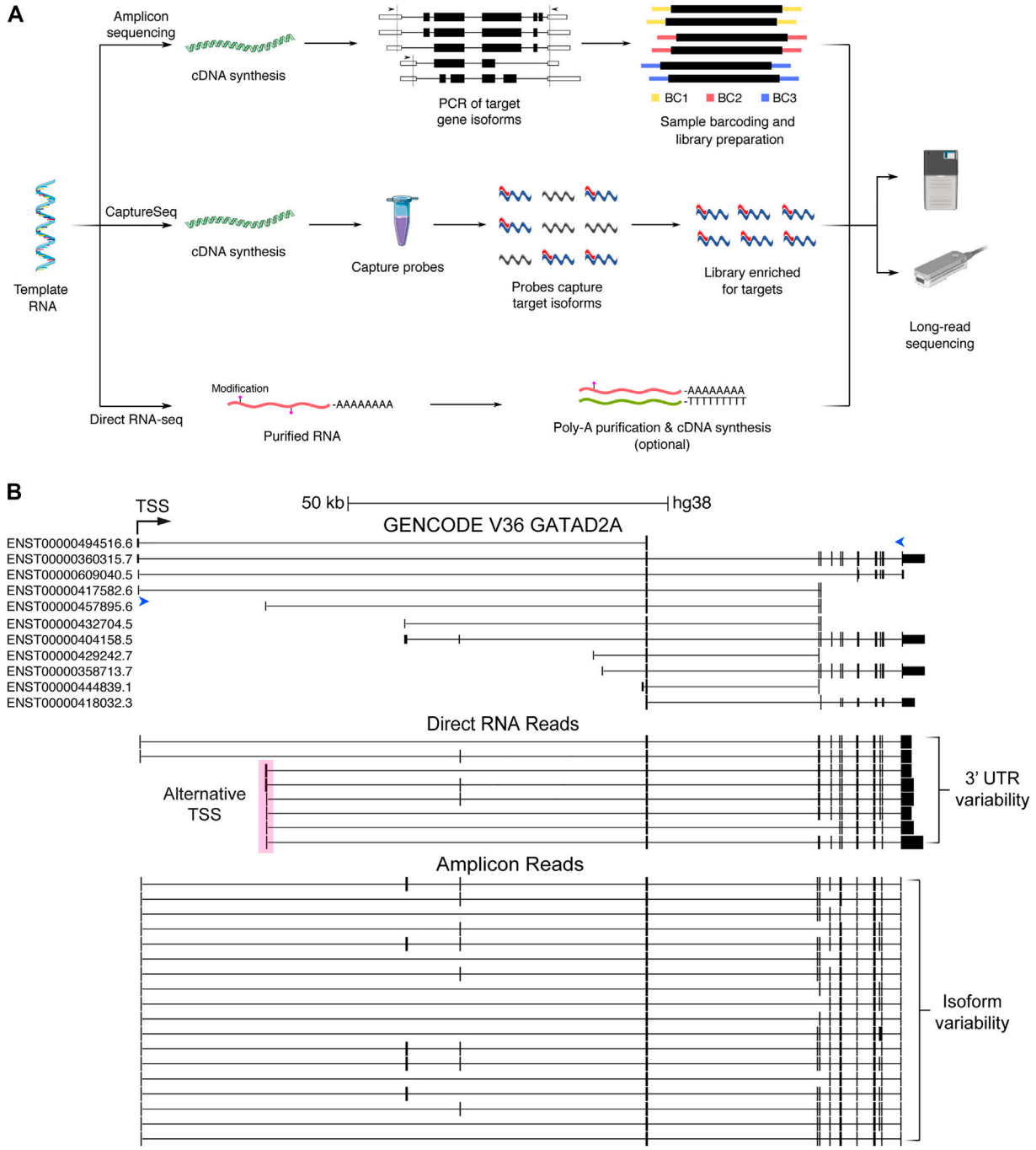

FIGURE 2 | Long-read sequencing methods and data generated. **(A)** Simplified pathways describing three long-read sequencing methods: amplicon sequencing, CaptureSeq and direct RNA sequencing. Amplicon sequencing involves cDNA synthesis followed by PCR of known and novel expressed isoforms from target genes (arrows indicate locations of forward and reverse primers). Sample barcodes can additionally be used for multiplexing before sequencing. CaptureSeq is performed on cDNA and utilises pools of oligonucleotide probes (capture probes). Probes (red) hybridise to isoforms from target genes (blue), which can then be purified, creating a sequencing library highly enriched for target isoforms. Direct RNA sequencing requires purified RNA and commonly includes the optional preparation steps to purify polyA RNA (to remove ribosomal RNA) and perform cDNA synthesis (to break-up RNA secondary structures) respectively. Only the RNA strand is subsequently sequenced. **(B)** Schematic of isoform information generated by whole-transcriptome direct RNA sequencing vs gene-targeted amplicon sequencing of cDNA. Image from the UCSC Genome Browser of the schizophrenia risk gene *GATAD2A* (Ripke et al., 2014), showing GENCODE annotations compared to nanopore direct RNA (middle) and amplicon (bottom) reads collapsed into high-confidence isoforms by FLAIR (Gleeson et al., 2020; Tang et al., 2020). Forward and reverse primers for amplicon sequencing are indicated by blue arrow heads. Transcriptional start site (TSS) shown by black arrow. An alternate start site (pink box) and 3′UTR variants are captured by direct RNA sequencing of the SH-SY5Y cell line. At least five supporting reads were required to identify high-confidence isoforms from direct RNA sequencing. A large number of additional isoforms are supported by amplicon sequencing using a more stringent threshold of 500 supporting reads. CDS, coding sequence; BC, barcode.

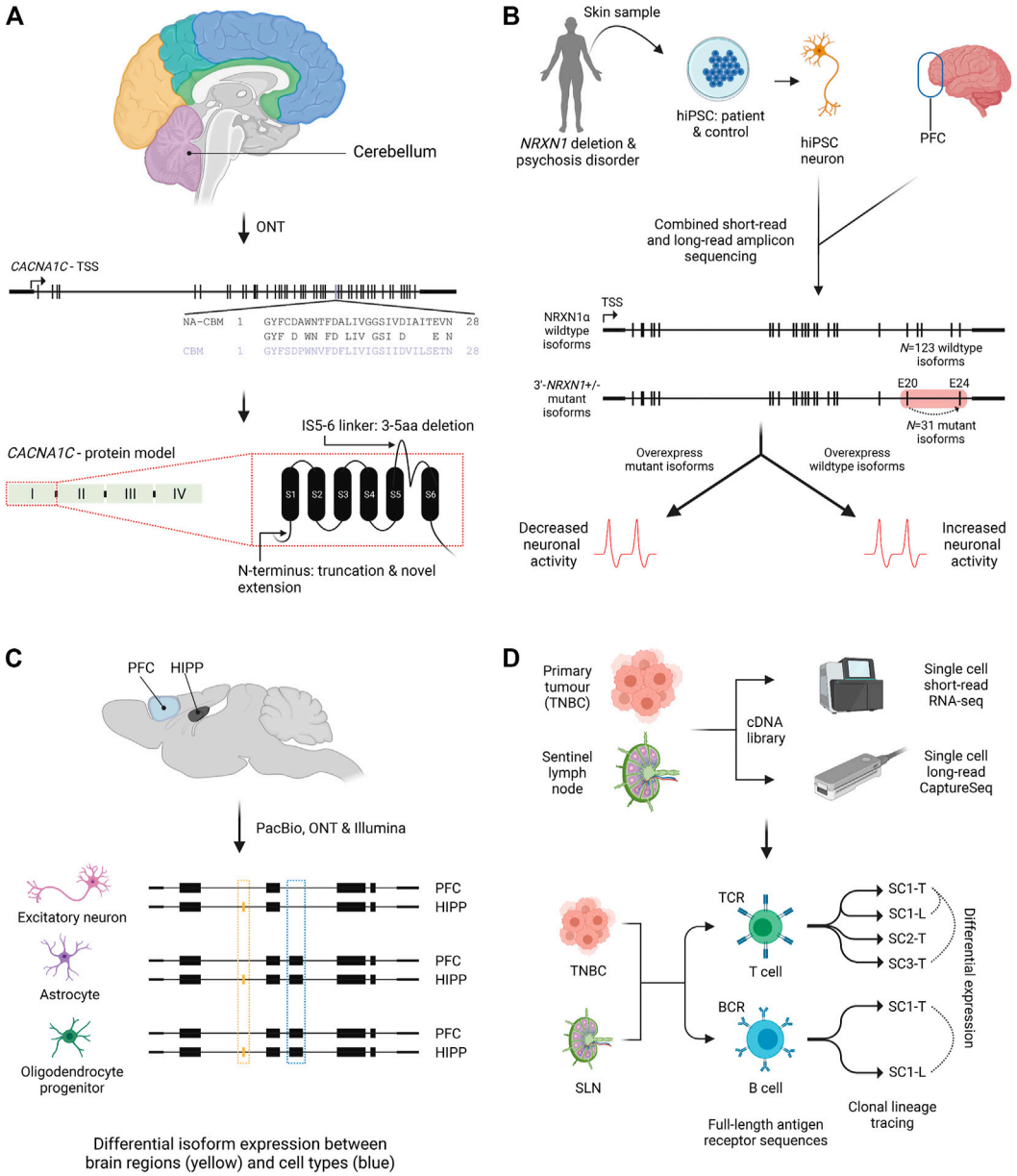

FIGURE 3 | Isoform characterisation in health and disease. **(A)** Nanopore amplicon sequencing of *CACNA1C* in post-mortem human brain found an isoform switch in cerebellum (purple) for exon 30 compared to non-cerebellar tissues. An example of novel splicing impacts on the N-terminus and an extracellular region (IS5-6 linker) is also shown for the first *CACNA1C* protein domain (Clark et al., 2020). NA-CBM, non-cerebellar tissue; CBM, cerebellum; aa, amino acid. **(B)** Combined short and long-read amplicon sequencing of *NRXN1α* in PFC and *NRNX1* mutant and wildtype hiPSCs characterised expressed isoforms. Overexpression of mutant isoforms in wildtype hiPSCs or wildtype isoforms in mutant iPSCs identified genotype dependent impacts on neuronal activity. Red box indicates location of patient specific *NRXN1α*$^{+/-}$ deletions. Dashed arrow indicates exon skipping from exons 20 to 24 (Flaherty et al., 2019). PFC, prefrontal cortex; hiPSC, human induced pluripotent stem cell; TSS, transcriptional start site. **(C)** Single cell isoform RNA sequencing (ScISOrSeq) from mouse prefrontal cortex and hippocampus identified cell type signatures, e.g., exon exclusion (blue box) and tissue specific signatures e.g. exon inclusion in hippocampus (orange box) (Joglekar et al., 2021). PFC, prefrontal cortex; HIPP, hippocampus. **(D)** Repertoire and Gene Expression by sequencing (RAGE-Seq) characterised full-length antigen receptor sequences for T and B cells from a primary tumour (breast) and its sentinel lymph node. Clonal lineages and clonal expansions were identified, as well as differential expression between clonally expanded T or B cells from paired tumour and lymph node samples (Singh et al., 2019). TNBC, triple negative breast cancer; SLN, sentinel lymph node; TCR, T cell receptor; BCR, B cell receptor; SC-T, shared clonotype tumour; SC-L, shared clonotype lymph node; ONT, Oxford Nanopore Technologies; PacBio, Pacific Biosciences.

generate capture probes quickly and cheaply from cDNA clones. OCS performed similarly to commercially synthesised probe pools and the authors demonstrated the flexible implementation of OCS on between 2 and 763 human transcription factors. PacBio sequencing of 763 targeted transcription factors identified a 7-fold increase in isoforms

compared to un-enriched samples. The development of OCS has the potential to broaden the applicability of CaptureSeq to a much wider range of studies (Sheynkman et al., 2020).

Generation of full-length isoforms using LCS takes a substantial step further towards a complete map of the human transcriptome, however there are some limitations that remain including the lower capture efficiency observed with long-reads and potential artifacts from incomplete reverse transcription, RNA degradation, or incorrectly identified splice junctions due to sequencing errors (Lagarde et al., 2017; Hardwick et al., 2019).

Long-Read Amplicon Sequencing

Unlike transcriptome-wide approaches or CaptureSeq, amplicon sequencing approaches are lower throughput and usually applied to a small number of genes. A schematic overview of the differences between target PCR, CaptureSeq and direct RNA sequencing is shown in **Figure 2A**, with examples of the different outputs from these shown in **Figure 2B**. A relatively straightforward but powerful approach is to use long-range PCR, with primers in the 5′ and 3′ UTRs, to amplify full-length isoforms (or entire coding regions) of genes of interest. The advantages of amplicon sequencing are extremely deep profiling of target genes and better coverage of longer isoforms, which are underrepresented when amplifying a pool of transcripts of varying lengths. Early PacBio long-read amplicon sequencing in bovine (*Bos taurus*) resulted in approximately 50,000 full-length immunoglobulin G (IgG) cDNA reads and the characterisation of a similar number of variable antigen binding regions, which was previously not possible with SRS (Larsen et al., 2012). Nanopore amplicon sequencing was subsequently used to identify almost 8,000 isoforms of *Dscam1*, the most extensively alternatively spliced gene known in *Drosophila* (Bolisetty et al., 2015).

Changes in splicing can be important in imparting risk for complex disease, however the isoform repertoire of many disease genes and which isoforms play a role in disease remains poorly understood (Li et al., 2016). The calcium channel *CACNA1C* has been identified by GWAS as a risk gene for neuropsychiatric disorders such as schizophrenia (Ripke et al., 2014). Amplicon sequencing from six human brain regions identified 38 novel exons and 241 novel *CACNA1C* isoforms (**Figure 3A**). Nine of the ten most abundant isoforms were novel, and many were predicted to encode channels with altered functions (Clark et al., 2020). Similarly, investigation of another schizophrenia risk gene, *SNX19*, in post-mortem human brain identified a group of alternatively spliced isoforms missing the *SNX19* C-terminal protein domain. Upregulation of these isoforms, several of which were novel, was associated with disease risk (Ma et al., 2020).

Neurexins are critical for synapse formation and studies have suggested hundreds of different isoforms may be produced, however efforts to identify these were hampered due to the length of the transcripts (~5 kb) (Ullrich et al., 1995). Treutlein et al. (2014) extensively profiled isoform diversity of three neurexin genes, *NRXN1*, *NRXN2* and *NRXN3* with PacBio amplicon sequencing in mouse prefrontal cortex (PFC). This unbiased approach identified between 9 and 258 isoforms per gene, the absence of any dominant isoform and the alternative splicing of an exon previously thought to be constitutive (Treutlein et al., 2014). A more recent hybrid approach performed both short and long-read amplicon sequencing to quantify *NRXN1α* isoforms in human PFC and in neurons derived from induced pluripotent stem cells (iPSCs) (Flaherty et al., 2019). A catalogue of 123 *NRXN1α* isoforms was identified, 86% of which showed conservation with the isoforms previously detected in mouse (**Figure 3B**). iPSC-neurons made from patients with *NRXN1α* mutations showed widespread isoform dysregulation, including loss of standard isoforms important for neuronal activity and expression of mutant isoforms inhibiting neuronal activity (Flaherty et al., 2019).

These studies on brain genes demonstrate that existing isoform annotations are far from complete and that novel transcripts may be playing a larger role in gene expression changes and in disease than previously understood. These findings can inform our understanding of the pathophysiology of complex neuropsychiatric disorders such as schizophrenia, where identifying the full repertoire of genes isoforms is a critical step towards a more genetically informed pathway to precision medicine (Treutlein et al., 2014; Flaherty et al., 2019). The identification of isoforms associated with disease risk, or disease-specific isoforms, also raises the possibility of creating isoform-specific therapeutics to target either the aberrant RNA or protein isoforms produced.

Long-read amplicon sequencing has also shown promise in contributing to the diagnosis of rare genetic diseases, especially those involving splicing changes. Long-read amplicon sequencing helped provide a molecular diagnosis for individuals with mitochondrial disease by confirming the inclusion of a cryptic exon in the mitochondrial Complex I subunit gene *NDUFB10*. This previously unannotated exon contained an early in-frame stop codon leading to nonsense mediated decay (NMD) of the altered transcript (Helman et al., 2021). This study highlighted the use of long-reads and multi-omic approaches to identify potential causes of disease where other genomic approaches have failed to determine a cause.

LONG-READ SEQUENCING FOR SINGLE CELL TRANSCRIPTOMICS

Single cell RNA sequencing (scRNA-Seq) profiles gene expression in individual cells. Short-read (SR) scRNA-Seq is now well established and has been very successful in identifying cell types and trajectories; cellular and tumour heterogeneity; expression differences between cell types and for performing high-throughput perturbation assays (Hwang et al., 2018). SR scRNA-Seq methods separate into two main types, transcript counting methods that sequence only the 3′ or 5′ ends of transcripts (such as the popular 10x Genomics platform) and whole transcript methods that sequence reads from all regions of an RNA [such as Smart-Seq2 (Picelli et al., 2013)]. Transcript counting methods can profile large numbers of cells but provide limited splicing or isoform information, while whole transcript methods are lower throughput but provide more information on

AS of exons. Single cell reads are typically tagged with unique molecular identifiers (UMIs - a unique sequence tag added to each molecule) and cell barcodes for accurate expression quantification and identification of the cell of origin respectively, information that needs to be measurable in any long-read (LR) scRNA-Seq method.

SR scRNA-Seq studies using "whole transcript" methods have identified significant cell-to-cell differences in isoform expression (Shalek et al., 2013; Marinov et al., 2014; Yap and Makeyev, 2016; Song et al., 2017). However, these studies largely focused on changes in the usage of specific exons or splice junctions due to short-read constraints, leaving the true complexity of AS and isoform expression within and between single cells unclear. The update to Smart-Seq3 (Hagemann-Jensen et al., 2020) incorporated UMIs into this methodology and improved isoform identification, however isoform reconstruction was poor beyond 1 kb and only ~40% of molecules could be assigned to an isoform. LR scRNA-Seq can be integrated with both 10x and Smart-Seq style methods by omitting cDNA fragmentation steps (Gupta et al., 2018; Singh et al., 2019). Coupling long-reads with single cell sequencing can provide the currently missing isoform information and has the potential to once again revolutionise transcriptomics.

Initial studies combining long reads with single cell technology were all performed on less than ten cells using either ONT (Byrne et al., 2017) or PacBio (Macaulay et al., 2015; Karlsson and Linnarsson, 2017) sequencing. While the number of cells was low, Byrne et al. (2017) found hundreds of genes expressing multiple isoforms and DIE, while Karlsson and Linnarsson. (2017) results suggested that isoform diversity was an important source of biological variability between cells. These studies demonstrated the potential of incorporating long-read sequencing into single cell studies, with a number of LR scRNA-Seq studies on the 10x chromium platform now emerging.

The first demonstration of LR scRNA-Seq in a significant number of cells utilised the PacBio LRS ScISOr-Seq method, complemented by SR scRNA-Seq, to study neurons, astrocytes and microglia from mouse cerebellum (Gupta et al., 2018). While isoforms from over one thousand single cells were characterised, ScISOr-Seq only sequenced a small median number of reads (270) and genes (129) per cell. The ScISOr-Seq method was built upon more recently to investigate cell-type specific splicing across mouse brain regions (**Figure 3C**) (Joglekar et al., 2021). Deeper per-cell profiling identified DIE in 395 genes between mouse hippocampus and prefrontal cortex cells, including 76 high-confidence novel isoforms. Of the 395 genes, 36% exhibited differential transcription start or end sites, while 64% had splice-site usage differences. DIE between brain regions was largely due to a single cell type changing its isoform expression pattern, a critical insight into the relative importance of cell-types, brain regions and cell composition in defining splicing patterns (Joglekar et al., 2021).

The higher error rate of nanopore sequencing provides additional challenges for LR scRNA-Seq due to the critical importance of identifying the cell barcode and UMIs present in each read. To address this, Lebrigand et al. (2020a) developed ScNaUmi-Seq where both LR and SR scRNA-Seq were performed on single cells from the 10x platform. SR scRNA-Seq identified the cell barcodes and UMIs present facilitating their identification in the higher error nanopore data. ScNaUmi-Seq was highly accurate, however cell barcodes and UMIs were only identified in ~30% of nanopore reads, suggesting significant room for future improvements through both increased read accuracy and enhanced identification algorithms. Applying ScNaUmi-Seq to over one thousand embryonic mouse brain cells identified a median of ~2,500 genes per cell and very high correlations between nanopore and SR gene counts, which validated that LR data gave an accurate representation of the transcriptome (Lebrigand et al., 2020a). Differential isoform usage (DIU) between cell-types was observed for 76 genes, including pronounced isoform switching during neuronal development (Lebrigand et al., 2020a).

As demonstrated by Gupta et al. (2018), sequencing large numbers of cells with current LR scRNA-Seq technologies either results in low per-cell read depth or high experimental cost. This creates a trade-off between per-cell depth (important for isoform comparisons) and number of cells sequenced. FLT-Seq (Tian et al., 2020) is a recently developed method that subsamples 10–20% of the cells from a 10x Genomics experiment for nanopore sequencing, which combined with SR scRNA-Seq on all cells allows for an integrated analysis using the FLAMES package (Dong et al., 2020). SR scRNA-Seq identifies the cell barcodes and UMIs present, (similar to ScNaUmi-Seq) while also providing a broader view of the cell-types present. LR scRNA-Seq on over two thousand human and mouse cells generated a number of insights into AS and isoform usage at the single cell level, including that the two most abundant isoforms most commonly differ by multiple AS events and that isoform proportions are largely unimodal, not bimodal, in a cell population (Tian et al., 2020).

The previously mentioned studies using LR scRNA-Seq relied on short-reads in addition to long-reads for tasks including increasing the number of profiled cells, cell clustering, improving the accuracy in assigning a long-read to its cell of origin and correcting long-read splice junctions. In contrast, the higher accuracy R2C2 method was recently used to sequence full-length nanopore reads and provide isoform-level data from ~1,500 blood cells without short-read support (Volden and Vollmers, 2020). The R2C2 method uses rolling circle amplification and concatemeric consensus sequencing to generate a read accuracy of 96% from nanopore data. SR sequencing was generated for comparison, confirming accurate gene quantification with R2C2. Along with establishing a short-read independent single cell method, a wide range of isoform diversity was found between genes, with high diversity genes like *LMNA* expressing a unique isoform in most cells (Volden and Vollmers, 2020). A trade-off of the R2C2 method is that by sequencing each cDNA multiple times for higher accuracy, the total number of unique cDNAs profiled is around 3x lower than standard methods. Additionally, 45% of R2C2 UMIs were unmatched in the SR data, demonstrating that even 96% accuracy is insufficient to assign many of the reads.

LR scRNA-Seq has also been combined with CaptureSeq to allow high sensitivity full-length profiling of isoforms of interest

at single cell resolution (Singh et al., 2019). T-cell and B-cell receptors (*TCRs* and *BCRs*) are incredibly diverse and unique to each T- or B- cell lineage due to DNA rearrangement, AS and somatic hypermutation, however typing and tracing lineages at the single cell level has proven difficult (Picelli et al., 2013; Afik et al., 2017; Rizzetto et al., 2018). RAGE-Seq uses targeted capture of *TCR* and *BCR* isoforms followed by nanopore LR scRNA-Seq, coupled with short-read expression profiling of single cells (Singh et al., 2019). Applied to 7,138 cells from a tumour and lymph node of a breast cancer patient (**Figure 3D**), lymphocyte clonotypes and clonal expansions could be identified as well as differential expression between different T-cell clonotypes. One limitation was the low assignment of cell barcodes due to the error rate of nanopore sequencing, however, the capture-based method can be applied to any transcripts of interest, potentially overcoming the low-depth observed in other studies such as Gupta et al. (2018).

While the methodology of combining single cell and long-read technologies is still in its infancy, improvements are being made at a rapid pace. In future, as throughput and/or accuracy improves, subsampling and matched SR scRNA-Seq may become unnecessary. Single cell direct RNA sequencing may also become possible, allowing for single molecule isoform expression and modification profiling. The currently available studies highlight the relevance of analysing isoforms at single cell resolution, and we anticipate future research in the field to provide novel insights into the splicing landscape of cells and tissues.

LONG-READ SEQUENCING FOR SPATIALLY RESOLVED TRANSCRIPTOMICS

While single cell sequencing has allowed us to uncover both known and novel cell types, the physical relationship between cells can now be studied with spatially resolved transcriptomics (Asp et al., 2020). Cells differentiate and function within tissues and are therefore influenced by their environment. Spatially resolved transcriptomics aims to profile gene expression patterns within this physical context. A large number of spatial methods have now been developed based on *in-situ* hybridisation, *in-situ* sequencing, or capture plus spatial barcoding of RNA (Asp et al., 2020).

The spatial transcriptomics method introduced by Ståhl et al. (2016) and now commercialised as the 10x Visium platform involves affixing tissue sections to slides covered with spatial barcodes allowing the generation of spatially barcoded cDNA. Two recent studies have adapted this platform for long-read sequencing. The first, spatial isoform transcriptomics (SiT), utilised nanopore long-reads combined with short-read sequencing for the assignment of UMIs and spatial barcodes (Lebrigand et al., 2020b). SiT was demonstrated in mouse brain, identifying 19 genes with isoform switching between regions in the olfactory bulb. For example, the *Plp1* gene involved in demyelination pathologies showed regional differences in isoform expression between the outer olfactory nerve layer and inner granule layer. In addition to spatial variation in isoform structure, SiT also identified spatial variation in isoform modifications. For example, A-to-I RNA editing events in the *Calm1* calcium receptor gene had robust variation between regions and showed a particularly high editing ratio in the thalamus (Lebrigand et al., 2020b). To investigate cell-type specific splicing and isoforms across mouse brain regions, Joglekar et al. (2021) developed slide-isoform sequencing (Sliso-Seq) to combine spatial transcriptomics with long-read sequencing. The author's previous ScISOr-Seq technique had identified DIE of *Snap25* as neuroblasts matured into excitatory neurons. Sliso-Seq subsequently showed the switch occurred in a posterior-to-anterior gradient across the brain, enabling the linking of single cell results to the spatial dynamics of the *Snap25* isoform switch (Joglekar et al., 2021).

Applying long-read technologies to spatial transcriptomics will deepen our understanding of how alternative isoform usage and splicing influence cell processes. Currently the 10x spatial transcriptomics platform adapted by Lebrigand et al. (2020b) and Joglekar et al. (2021) does not provide single cell resolution. However, in the future, coupling long-read and single cell spatial transcriptomics methods will allow for the creation of three-dimensional maps of isoform expression from single cells. These insights will provide valuable knowledge into developmental mechanisms and pathways involved in disease.

REMAINING CHALLENGES FOR LONG-READ SEQUENCING

Long-read sequencing approaches excel at confirming exon connectivity and general transcript structure, however there are several limitations that users should be aware of when interpreting results. Amplification of target genes or whole transcriptomes carries the risk of PCR bias (shorter transcripts are favoured by PCR and will increase in proportion as more cycles are performed) and generation of chimeric cDNA (Brakenhoff et al., 1991; Bolisetty et al., 2015). Both issues can be mitigated by minimising PCR cycles, performing direct cDNA or RNA sequencing or through the use of UMIs (Lebrigand et al., 2020a; Karst et al., 2021). UMIs proved particularly useful for identifying robust differential isoform usage (DIU) in genes that appeared to have consistent expression between conditions (Lebrigand et al., 2020a). UMIs can also be used to reduce the relatively high error rates from nanopore sequencing by clustering together multiple reads from an original sequence (Karst et al., 2021).

Sample quality is critical for long-read data as degraded RNA or DNA can lead to fragmented sequences negating the primary advantage of long-read approaches. The quality of RNA, measured by the RNA integrity number (RIN), can vary significantly between samples and sample types. A minimum RIN of 6 (optimally > 7), was recommended for long-read amplicon sequencing (Clark et al., 2020), however the necessary RNA integrity for unbiased transcriptome-wide sequencing could potentially be higher. Additionally, post-mortem tissues are commonly used for human transcriptome analysis and the impact of death, post-mortem interval and ischemia should be considered when interpreting gene expression data (Ferreira et al., 2018).

The accuracy of ONT and PacBio long reads is generally lower than that of SRS and this can impact determination of splice sites and isoform identity as well as barcode and UMI assignment (Amarasinghe et al., 2020). In addition to the error rate, long-read methods do not always generate full-length reads due to compromised sample quality, library preparation limitations and incompletely sequenced reads. This further complicates isoform identification and quantification as reads can have multiple transcriptome alignments. Therefore, correctly identifying the isoform of origin is still a challenging task (Sessegolo et al., 2019; Soneson et al., 2019). Error correction and isoform identification tools, such as SQANTI and FLAIR, correct reads using long-read consensus or paired SRS data (Tardaguila et al., 2018; Tang et al., 2020). These tools address truncated reads by collapsing partial isoforms into the longer isoforms they likely represent and setting a minimum read threshold for an isoform to be considered valid. Benchmarking of these tools and the details of different error correction approaches and novel isoform assignment has been reviewed in detail elsewhere (Zhang et al., 2017; Amarasinghe et al., 2020), however false positive isoform identification remains a challenge. While software improvements are an avenue for improvement, isoform identification will also improve with decreased errors rates and sample preparation techniques that lead to less truncated reads (Amarasinghe et al., 2020).

THE IMPACT OF ALTERNATIVE SPLICING ON PROTEIN FUNCTION AND ABUNDANCE

As long-read sequencing continues to build a foundation of high-confidence isoforms, particularly for complex genes, the question of how these isoform changes functionally impact the proteome can begin to be answered. There is considerable debate over functional roles for alternative transcripts, with some studies reporting a single major isoform for most genes and little evidence for protein variants due to alternative splicing (Djebali et al., 2012; Gonzàlez-Porta et al., 2013; Tress et al., 2017). Other studies have shown evidence to the contrary, including the Vertebrate Alternative Splicing and Transcription Database (VastDB), which reported pronounced isoform switching in 48% of multi-exonic human genes and multiple co-expressed major isoforms in 18.5% (Tapial et al., 2017). Up to 75% of isoforms with exon-skipping have also been reported to be engaged by the ribosome, suggesting many are translated (Weatheritt et al., 2016). In addition, a splicing perturbation study by Liu et al. (2017) demonstrated that changes in RNA isoform expression led to changes in protein expression. More broadly, alternative isoforms such as non-coding isoforms, NMD isoforms or isoforms with altered UTR sequences can regulate gene expression and function even if the proteins produced do not change. Some of the debate may be down to perspective, many genes do appear to have a single dominant isoform in a cell type, however this does not mean other isoforms are not functionally relevant or become more prominent in other tissues or developmental timepoints, for which there is also plentiful evidence (Lebrigand et al., 2020a; Lebrigand et al., 2020b). In contrast, other genes have extremely complex isoform expression patterns and are poorly annotated by prior methods (Clark et al., 2020). Therefore, our understanding of the functional impact of alternative isoforms is constantly evolving (Weatheritt et al., 2016; Ule and Blencowe, 2019).

In their targeted study of the calcium channel gene *CACNA1C*, Clark et al. (2020) predicted the impact of novel full-length RNA isoforms on protein function. They found 51/83 (~61%) of novel high-confidence isoforms detected using nanopore sequencing potentially encoded a functional Ca^{2+} channel. An additional splice-site-level analysis identified almost half of the isoforms had splice junction changes encoding microdeletions of 3–5 amino acids within regions previously implicated in channel conductance (Clark et al., 2020). Such results will be important to follow up. Several computational approaches are available to predict the impact of alternative RNA isoforms on protein features. For example, protein 3D structure and solubility can be predicted with programs such as Alternative Splicing-induced ALteration of Protein Structure (AS-ALPS) (Shionyu et al., 2009), DeepGOPlus (Kulmanov and Hoehndorf, 2019), PaRSnIP (Rawi et al., 2018), GOLabeler (You et al., 2018) and Deep Splicing Code (DSC) (Louadi et al., 2019).

Whilst computational predictive and simulation-based approaches are useful, validation using established and emerging proteomic technologies is the gold standard for assessing any biological impact of alternative mRNA isoforms on the proteome. Like RNA sequencing, proteomic methods have their caveats, identifying novel sequences and protein isoforms is generally more difficult with mass spectrometry than for transcriptomics (Blencowe, 2017). In addition, understanding the relative detection sensitivity of each method will be essential for successfully quantifying the relationship between RNA and protein isoforms. Thanks to recent advances, integration of genomics, transcriptomics and proteomics to improve our understanding of traits and diseases is now feasible, opening up new opportunities to examine the role of RNA and protein isoforms in health and disease (Molendijk and Parker, 2021).

DISCUSSION

Long-read sequencing enables profiling of full-length RNA and cDNA reads, which is essential for mapping alternative RNA isoforms in tissues and disease states. Coupling long-read data with both short reads and cutting-edge technologies such as single cell sequencing significantly widens the toolset for accurate isoform discovery in complex transcriptomes. Long read methods currently involve a trade-off between higher accuracy (Hifi, R2C2) and higher throughput (Nanopore, PacBio subreads). While lower accuracy can necessitate sophisticated error correction tools and/or paired short-read data, the advantages of long-over short-reads for isoform detection and quantification already outweigh many of the drawbacks in error rate (Chen et al., 2021). Furthermore, we anticipate that error correction will not be necessary in future due to the rapid pace at which long-read technologies are improving.

Long-read transcriptomics is still a nascent field of research, however it has already had a major impact on our understanding

of spliced isoform diversity and expression (Lebrigand et al., 2020a; Clark et al., 2020; Joglekar et al., 2021). As transcriptomics moves into the age of the isoform, long-read technologies that enable transcriptomic characterisation at single cell and spatial resolutions will lead the way for new discoveries in health, development and disease.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by an Australian National Health and Medical Research Council Investigator Grant (APP1196841) to MC.

ACKNOWLEDGMENTS

We would like to thank Yair Prawer and Shweta Joshi for their helpful feedback. **Figures 2**, **3** were created under license with BioRender.com.

REFERENCES

Afik, S., Yates, K. B., Bi, K., Darko, S., Godec, J., Gerdemann, U., et al. (2017). Targeted Reconstruction of T Cell Receptor Sequence from Single Cell RNA-Seq Links CDR3 Length to T Cell Differentiation State. *Nucleic Acids Res.* 45, e148. doi:10.1093/nar/gkx615

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., and Gouil, Q. (2020). Opportunities and Challenges in Long-Read Sequencing Data Analysis. *Genome Biol.* 21, 30–16. doi:10.1186/s13059-020-1935-5

Ambardar, S., Gupta, R., Trakroo, D., Lal, R., and Vakhlu, J. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J. Microbiol.* 56, 394–404. doi:10.1007/s12088-016-0606-4

Anvar, S. Y., Allard, G., Tseng, E., Sheynkman, G. M., de Klerk, E., Vermaat, M., et al. (2018). Full-length mRNA Sequencing Uncovers a Widespread Coupling between Transcription Initiation and mRNA Processing. *Genome Biol.* 19, 46. doi:10.1186/s13059-018-1418-0

Asnani, M., Hayer, K. E., Naqvi, A. S., Zheng, S., Yang, S. Y., Oldridge, D., et al. (2020). Retention of CD19 Intron 2 Contributes to CART-19 Resistance in Leukemias with Subclonal Frameshift Mutations in CD19. *Leukemia* 34, 1202–1207. doi:10.1038/s41375-019-0580-z

Asp, M., Bergenstråhle, J., and Lundeberg, J. (2020). Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *BioEssays* 42, 1900221. doi:10.1002/bies.201900221

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., et al. (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science* 338, 1587–1593. doi:10.1126/science.1230612

Bartonicek, N., Clark, M. B., Quek, X. C., Torpy, J. R., Pritchard, A. L., Maag, J. L. V., et al. (2017). Intergenic Disease-Associated Regions Are Abundant in Novel Transcripts. *Genome Biol.* 18, 241. doi:10.1186/s13059-017-1363-3

Bhatt, D. M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M. M., Natoli, G., et al. (2012). Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. *Cell* 150, 279–290. doi:10.1016/j.cell.2012.05.043

Blencowe, B. J. (2017). The Relationship between Alternative Splicing and Proteomic Complexity. *Trends Biochem. Sci.* 42, 407–408. doi:10.1016/j.tibs.2017.04.001

Bolisetty, M. T., Rajadinakaran, G., and Graveley, B. R. (2015). Determining Exon Connectivity in Complex mRNAs by Nanopore Sequencing. *Genome Biol.* 16, 204. doi:10.1186/s13059-015-0777-z

Brakenhoff, R. H., Schoenmakers, J. G. G., Lubsen, N. H., Brakenhoff, R., Schoenmakers, G. G., Lubsen, J. H., et al. (1991). Chimeric cDNA Clones: A Novel PCR Artifact. *Nucl. Acids Res.* 19, 1949. doi:10.1093/nar/19.8.1949

Braspenning, S. E., Sadaoka, T., Breuer, J., Verjans, G. M. G. M., Ouwendijk, W. J. D., and Depledge, D. P. (2020). Decoding the Architecture of the Varicella-Zoster Virus Transcriptome. *mBio* 11. doi:10.1128/mBio.01568-20

Buniello, A., Macarthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of Published Genome-wide Association Studies, Targeted Arrays and Summary Statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi:10.1093/nar/gky1120

Byrne, A., Beaudin, A. E., Olsen, H. E., Jain, M., Cole, C., Palmer, T., et al. (2017). Nanopore Long-Read RNAseq Reveals Widespread Transcriptional Variation Among the Surface Receptors of Individual B Cells. *Nat. Commun.* 8, 16027. doi:10.1038/ncomms16027

Byrne, A., Cole, C., Volden, R., and Vollmers, C. (2019). Realizing the Potential of Full-Length Transcriptome Sequencing. *Phil. Trans. R. Soc. B* 374, 1786. doi:10.1098/rstb/374/1786

Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., et al. (1996). High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* 37, 327–336. doi:10.1006/geno.1996.0567

Chang, J. J.-Y., Rawlinson, D., Pitt, M. E., Taiaroa, G., Gleeson, J., Zhou, C., et al. (2021). Transcriptional and Epi-Transcriptional Dynamics of SARS-CoV-2 during Cellular Infection. *Cell Rep.* 35. doi:10.1016/j.celrep.2021.109108

Chen, H., Gao, F., He, M., Ding, X. F., Wong, A. M., Sze, S. C., et al. (2019). Long-Read RNA Sequencing Identifies Alternative Splice Variants in Hepatocellular Carcinoma and Tumor-Specific Isoforms. *Hepatology* 70, 1011–1025. doi:10.1002/hep.30500

Chen, S. Y., Deng, F., Jia, X., Li, C., and Lai, S. J. (2017). A Transcriptome Atlas of Rabbit Revealed by PacBio Single-Molecule Long-Read Sequencing. *Sci. Rep.* 7, 7648. doi:10.1038/s41598-017-08138-z

Chen, Y., Davidson, N. M., Wan, Y. K., Patel, H., Yao, F., Low, H. M., et al. (2021). *A Systematic Benchmark of Nanopore Long Read RNA Sequencing for Transcript Level Analysis in Human Cell Lines*. bioRxiv, 440736. doi:10.1101/2021.04.21.440736

Clark, M. B., Mercer, T. R., Bussotti, G., Leonardi, T., Haynes, K. R., Crawford, J., et al. (2015). Quantitative Gene Profiling of Long Noncoding RNAs with Targeted RNA Sequencing. *Nat. Methods* 12, 339–342. doi:10.1038/nmeth.3321

Clark, M. B., Wrzesinski, T., Garcia, A. B., Hall, N. A. L., Kleinman, J. E., Hyde, T., et al. (2020). Long-read Sequencing Reveals the Complex Splicing Profile of the Psychiatric Risk Gene CACNA1C in Human Brain. *Mol. Psychiatry* 25, 37–47. doi:10.1038/s41380-019-0583-1

Deamer, D., Akeson, M., and Branton, D. (2016). Three Decades of Nanopore Sequencing. *Nat. Biotechnol.* 34, 518–524. doi:10.1038/nbt.3423

Deveson, I. W., Brunck, M. E., Blackburn, J., Tseng, E., Hon, T., Clark, T. A., et al. (2018). Universal Alternative Splicing of Noncoding Exons. *Cell Syst.* 6, 245–255. doi:10.1016/j.cels.2017.12.005

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of Transcription in Human Cells. *Nature* 489, 101–108. doi:10.1038/nature11233

Dong, X., Tian, L., Gouil, Q., Kariyawasam, H., Su, S., Paoli-Iseppi, R. D., et al. (2020). The Long and the Short of it: Unlocking Nanopore Long-Read RNA Sequencing Data with Short-Read Differential Expression Analysis Tools. *NAR Genomics and Bioinformatics* 3, doi:10.1093/nargab/lqab028

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA Sequencing from Single Polymerase Molecules. *Science* 323, 133–138. doi:10.1126/science.1162986

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., et al. (2008). Genetics of Gene Expression and its Effect on Disease. *Nature* 452, 423–428. doi:10.1038/nature06758

Engström, P. G., Steijger, T., Steijger, T., Sipos, B., Grant, G. R., Kahles, A., et al. (2013). Systematic Evaluation of Spliced Alignment Programs for RNA-Seq Data. *Nat. Methods* 10, 1185–1191. doi:10.1038/nmeth.2722

Ferreira, P. G., Muñoz-Aguirre, M., Reverter, F., Sá Godinho, C. P., Sousa, A., Amadoz, A., et al. (2018). The Effects of Death and post-mortem Cold Ischemia on Human Tissue Transcriptomes. *Nat. Commun.* 9, 490. doi:10.1038/s41467-017-02772-x

Flaherty, E., Zhu, S., Barretto, N., Cheng, E., Deans, P. J. M., Fernando, M. B., et al. (2019). Neuronal Impact of Patient-specific Aberrant NRXN1α Splicing. *Nat. Genet.* 51, 1679–1690. doi:10.1038/s41588-019-0539-z

Fujiyoshi, S., Muto-Fujita, A., and Maruyama, F. (2020). Evaluation of PCR Conditions for Characterizing Bacterial Communities with Full-Length 16S rRNA Genes Using a Portable Nanopore Sequencer. *Sci. Rep.* 10, 12580. doi:10.1038/s41598-020-69450-9

Furney, S. J., Pedersen, M., Gentien, D., Dumont, A. G., Rapinat, A., Desjardins, L., et al. (2013). SF3B1 Mutations Are Associated with Alternative Splicing in Uveal Melanoma. *Cancer Discov.* 3, 1122–1129. doi:10.1158/2159-8290.cd-13-0330

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., et al. (2018). Highly Parallel Direct RNA Sequencing on an Array of Nanopores. *Nat. Methods* 15, 201–206. doi:10.1038/nmeth.4577

Gleeson, J., Lane, T. A., Harrison, P. J., Haerty, W., and Clark, M. B. (2020). *Nanopore Direct RNA Sequencing Detects Differential Expression between Human Cell Populations*. bioRxiv, 232785. doi:10.1101/2020.08.02.232785

Glinos, D. A., Garborcauskas, G., Hoffman, P., Ehsan, N., Jiang, L., Gokden, A., et al. (2021). *Transcriptome Variation in Human Tissues Revealed by Long-Read Sequencing*. bioRxiv, 427687. doi:10.1101/2021.01.22.427687

Gonzalez-Garay, M. L. (2016). "Introduction to Isoform Sequencing Using Pacific Biosciences Technology (Iso-Seq)," in *Transcriptomics and Gene Regulation* (Springer), 9, 141–160. doi:10.1007/978-94-017-7450-5_6

Gonzàlez-Porta, M., Frankish, A., Rung, J., Harrow, J., and Brazma, A. (2013). Transcriptome Analysis of Human Tissues and Cell Lines Reveals One Dominant Transcript Per Gene. *Genome Biol.* 14, R70. doi:10.1186/gb-2013-14-7-r70

Gupta, I., Collier, P. G., Haase, B., Mahfouz, A., Joglekar, A., Floyd, T., et al. (2018). Single-cell Isoform RNA Sequencing Characterizes Isoforms in Thousands of Cerebellar Cells. *Nat. Biotechnol.* 36, 1197–1202. doi:10.1038/nbt.4259

Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G.-J., Larsson, A. J. M., et al. (2020). Single-cell RNA Counting at Allele and Isoform Resolution Using Smart-Seq3. *Nat. Biotechnol.* 38, 708–714. doi:10.1038/s41587-020-0497-0

Hardwick, S. A., Bassett, S. D., Kaczorowski, D., Blackburn, J., Barton, K., Bartonicek, N., et al. (2019). Targeted, High-Resolution RNA Sequencing of Non-coding Genomic Regions Associated with Neuropsychiatric Functions. *Front. Genet.* 10, 309. doi:10.3389/fgene.2019.00309

Helman, G., Compton, A. G., Hock, D. H., Walkiewicz, M., Brett, G. R., Pais, L., et al. (2021). Multiomic Analysis Elucidates Complex I Deficiency Caused by a Deep Intronic Variant in NDUFB10. *Hum. Mutat.* 42, 19–24. doi:10.1002/humu.24135

Hon, T., Mars, K., Young, G., Tsai, Y.-C., Karalius, J. W., Landolin, J. M., et al. (2020). Highly Accurate Long-Read HiFi Sequencing Data for Five Complex Genomes. *Sci. Data* 7, 399. doi:10.1038/s41597-020-00743-4

Huang, K. K., Huang, J., Wu, J. K. L., Lee, M., Tay, S. T., Kumar, V., et al. (2021). Long-read Transcriptome Sequencing Reveals Abundant Promoter Diversity in Distinct Molecular Subtypes of Gastric Cancer. *Genome Biol.* 22, 1–24. doi:10.1186/s13059-021-02261-x

Hwang, B., Lee, J. H., and Bang, D. (2018). Single-cell RNA Sequencing Technologies and Bioinformatics Pipelines. *Exp. Mol. Med.* 50, 1–14. doi:10.1038/s12276-018-0071-8

Jain, M., Olsen, H. E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community. *Genome Biol.* 17, 239. doi:10.1186/s13059-016-1103-0

Joglekar, A., Prjibelski, A., Mahfouz, A., Collier, P., Lin, S., Schlusche, A. K., et al. (2021). A Spatially Resolved Brain Region- and Cell Type-specific Isoform Atlas of the Postnatal Mouse Brain. *Nat. Commun.* 12, 463. doi:10.1038/s41467-020-20343-5

Johnson, R. W., Rice, A. S., and Rice, A. S. C. (2014). Clinical Practice. Postherpetic Neuralgia. *N. Engl. J. Med.* 371, 1526–1533. doi:10.1056/NEJMcp1403062

Kahles, A., Lehmann, K. V., Toussaint, N., Hüser, M., Stark, S. G., Sachsenberg, T., et al. (2018). Comprehensive Analysis of Alternative Splicing across Tumors from 8,705 Patients. *Cancer Cell* 34, 211–e6. doi:10.1016/j.ccell.2018.07.001

Kahraman, A., Karakulak, T., Szklarczyk, D., and von Mering, C. (2020). Pathogenic Impact of Transcript Isoform Switching in 1,209 Cancer Samples Covering 27 Cancer Types Using an Isoform-specific Interaction Network. *Sci. Rep.* 10, 14453. doi:10.1038/s41598-020-71221-5

Karlsson, K., and Linnarsson, S. (2017). Single-cell mRNA Isoform Diversity in the Mouse Brain. *BMC Genomics* 18, 126. doi:10.1186/s12864-017-3528-6

Karst, S. M., Ziels, R. M., Kirkegaard, R. H., Sørensen, E. A., McDonald, D., Zhu, Q., et al. (2021). High-accuracy Long-Read Amplicon Sequences Using Unique Molecular Identifiers with Nanopore or PacBio Sequencing. *Nat. Methods* 18, 165–169. doi:10.1038/s41592-020-01041-y

Kim, J.-A., Roy, N. S., Lee, I.-h., Choi, A.-Y., Choi, B.-S., Yu, Y.-S., et al. (2019). Genome-wide Transcriptome Profiling of the Medicinal Plant Zanthoxylum Planispinum Using a Single-Molecule Direct RNA Sequencing Approach. *Genomics* 111, 973–979. doi:10.1016/j.ygeno.2018.06.004

Kulmanov, M., and Hoehndorf, R. (2019). DeepGOPlus: Improved Protein Function Prediction from Sequence. *Bioinformatics* 36, 422–429. doi:10.1093/bioinformatics/btz595

Lagarde, J., Uszczynska-Ratajczak, B., Carbonell, S., Pérez-Lluch, S., Abad, A., Davis, C., et al. (2017). High-throughput Annotation of Full-Length Long Noncoding RNAs with Capture Long-Read Sequencing. *Nat. Genet.* 49, 1731–1740. doi:10.1038/ng.3988

Larsen, P. A., Smith, T. P. L., Larsen, A., and Smith, T. (2012). Application of Circular Consensus Sequencing and Network Analysis to Characterize the Bovine IgG Repertoire. *BMC Immunol.* 13, 52. doi:10.1186/1471-2172-13-52

Lebrigand, K., Bergenstråhle, J., Thrane, K., Mollbrink, A., Barbry, P., Waldmann, R., et al. (2020b). The Spatial Landscape of Gene Expression Isoforms in Tissue Sections. *bioRxiv*, 252296. doi:10.1101/2020.08.24.252296

Lebrigand, K., Magnone, V., Barbry, P., and Waldmann, R. (2020a). High Throughput Error Corrected Nanopore Single Cell Transcriptome Sequencing. *Nat. Commun.* 11, 4025. doi:10.1038/s41467-020-17800-6

Lee, T. I., and Young, R. A. (2013). Transcriptional Regulation and its Misregulation in Disease. *Cell* 152, 1237–1251. doi:10.1016/j.cell.2013.02.014

Li, Y. I., van de Geijn, B., Raj, A., Knowles, D. A., Petti, A. A., Golan, D., et al. (2016). RNA Splicing Is a Primary Link between Genetic Variation and Disease. *Science* 352, 600–604. doi:10.1126/science.aad9417

Lian, B., Hu, X., and Shao, Z. M. (2019). Unveiling Novel Targets of Paclitaxel Resistance by Single Molecule Long-Read RNA Sequencing in Breast Cancer. *Sci. Rep.* 9, 6032. doi:10.1038/s41598-019-42184-z

Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J., and Fairbrother, W. G. (2011). Using Positional Distribution to Identify Splicing Elements and Predict Pre-mRNA Processing Defects in Human Genes. *Proc. Natl. Acad. Sci.* 108, 11093–11098. doi:10.1073/pnas.1101135108

Liu, Y., Gonzàlez-Porta, M., Santos, S., Brazma, A., Marioni, J. C., Aebersold, R., et al. (2017). Impact of Alternative Splicing on the Human Proteome. *Cell Rep.* 20, 1229–1241. doi:10.1016/j.celrep.2017.07.025

Lorenz, D. A., Sathe, S., Einstein, J. M., Yeo, G. W., A Lorenz, D., Sathe, S., et al. (2020). Direct RNA Sequencing Enables m6A Detection in Endogenous Transcript Isoforms at Base-specific Resolution. *RNA* 26, 19–28. doi:10.1261/rna.072785.119

Louadi, Z., Oubounyt, M., Tayara, H., and Chong, K. T. (2019). Deep Splicing Code: Classifying Alternative Splicing Events Using Deep Learning. *Genes* 10, 587. doi:10.3390/genes10080587

Ma, L., Semick, S. A., Semick, S. A., Chen, Q., Li, C., Tao, R., et al. (2020). Schizophrenia Risk Variants Influence Multiple Classes of Transcripts of Sorting Nexin 19 (SNX19). *Mol. Psychiatry* 25, 831–843. doi:10.1038/s41380-018-0293-0

Macaulay, I. C., Haerty, W., Kumar, P., Li, Y. I., Hu, T. X., Teng, M. J., et al. (2015). G&T-seq: Parallel Sequencing of Single-Cell Genomes and Transcriptomes. *Nat. Methods* 12, 519–522. doi:10.1038/nmeth.3370

Marinov, G. K., Williams, B. A., McCue, K., Schroth, G. P., Gertz, J., Myers, R. M., et al. (2014). From Single-Cell to Cell-Pool Transcriptomes: Stochasticity in Gene Expression and RNA Splicing. *Genome Res.* 24, 496–510. doi:10.1101/gr.161034.113

Mercer, T. R., Clark, M. B., Crawford, J., Brunck, M. E., Gerhardt, D. J., Taft, R. J., et al. (2014). Targeted Sequencing for Gene Discovery and Quantification Using RNA CaptureSeq. *Nat. Protoc.* 9, 989–1009. doi:10.1038/nprot.2014.058

Molendijk, J., and Parker, B. L. (2021). Proteome-wide Systems Genetics to Identify Functional Regulators of Complex Traits. *Cell Syst.* 12, 5–22. doi:10.1016/j.cels.2020.10.005

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi:10.1038/nmeth.1226

Nilsen, T. W., and Graveley, B. R. (2010). Expansion of the Eukaryotic Proteome by Alternative Splicing. *Nature* 463, 457–463. doi:10.1038/nature08909

Oikonomopoulos, S., Wang, Y. C., Djambazian, H., Badescu, D., and Ragoussis, J. (2016). Benchmarking of the Oxford Nanopore MinION Sequencing for Quantitative and Qualitative Assessment of cDNA Populations. *Sci. Rep.* 6, 31602. doi:10.1038/srep31602

Oka, M., Xu, L., Suzuki, T., Yoshikawa, T., Sakamoto, H., Uemura, H., et al. (2021). Aberrant Splicing Isoforms Detected by Full-Length Transcriptome Sequencing as Transcripts of Potential Neoantigens in Non-small Cell Lung Cancer. *Genome Biol.* 22, 9–30. doi:10.1186/s13059-020-02240-8

Oxford Nanopore Technologies (2021). Guppy v5.0.7 Release. https://community.nanoporetech.com/posts/guppy-v5-0-7-release-note.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing. *Nat. Genet.* 40, 1413–1415. doi:10.1038/ng.259

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells. *Nat. Methods* 10, 1096–1098. doi:10.1038/nmeth.2639

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., et al. (2015). Rapid Draft Sequencing and Real-Time Nanopore Sequencing in a Hospital Outbreak of Salmonella. *Genome Biol.* 16, 114. doi:10.1186/s13059-015-0677-2

Rawi, R., Mall, R., Kunji, K., Shen, C.-H., Kwong, P. D., and Chuang, G.-Y. (2018). PaRSnIP: Sequence-Based Protein Solubility Prediction Using Gradient Boosting Machine. *Bioinformatics* 34, 1092–1098. doi:10.1093/bioinformatics/btx662

Rhine, C. L., Cygan, K. J., Soemedi, R., Maguire, S., Murray, M. F., Monaghan, S. F., et al. (2018). Hereditary Cancer Genes Are Highly Susceptible to Splicing Mutations. *Plos Genet.* 14, e1007231. doi:10.1371/journal.pgen.1007231

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K.-H. H., Holmans, P. A., et al. (2014). Biological Insights from 108 Schizophrenia-Associated Genetic Loci. *Nature* 511, 421–427. doi:10.1038/nature13595

Rizzetto, S., Koppstein, D. N. P., Samir, J., Singh, M., Reed, J. H., Cai, C. H., et al. (2018). B-cell Receptor Reconstruction from Single-Cell RNA-Seq with VDJPuzzle. *Bioinformatics* 34, 2846–2847. doi:10.1093/bioinformatics/bty203

Roach, N. P., Sadowski, N., Alessi, A. F., Timp, W., Taylor, J., and Kim, J. K. (2020). The Full-Length Transcriptome of *C. elegans* Using Direct RNA Sequencing. *Genome Res.* 30, 299–312. doi:10.1101/gr.251314.119

Robinson, E. K., Jagannatha, P., Covarrubias, S., Cattle, M., Safavi, R., Song, R., et al. (2020). *Inflammation Drives Alternative First Exon Usage to Regulate Immune Genes Including a Novel Iron Regulated Isoform of Aim2*. bioRxiv, 190330. doi:10.1101/2020.07.06.190330

Roundtree, I. A., and He, C. (2016). RNA Epigenetics - Chemical Messages for Posttranscriptional Gene Regulation. *Curr. Opin. Chem. Biol.* 30, 46–51. doi:10.1016/j.cbpa.2015.10.024

Russell, J. A., Campos, B., Stone, J., Blosser, E. M., Burkett-Cadena, N., Jacobs, J. L., et al. (2018). Unbiased Strain-Typing of Arbovirus Directly from Mosquitoes Using Nanopore Sequencing: A Field-Forward Biosurveillance Protocol. *Sci. Rep.* 8, 5417. doi:10.1038/s41598-018-23641-7

Sahlin, K., Medvedev, P., James, P. L., and Medvedev, P. (2021). Error Correction Enables Use of Oxford Nanopore Technology for Reference-free Transcriptome Analysis. *Nat. Commun.* 12, 2. doi:10.1038/s41467-020-20340-8

Sciarrillo, R., Wojtuszkiewicz, A., Assaraf, Y. G., Jansen, G., Kaspers, G. J. L., Giovannetti, E., et al. (2020). The Role of Alternative Splicing in Cancer: From Oncogenesis to Drug Resistance. *Drug Resist. Updates* 53, 100728. doi:10.1016/j.drup.2020.100728

Seki, M., Katsumata, E., Suzuki, A., Sereewattanawoot, S., Sakamoto, Y., Mizushima-Sugano, J., et al. (2019). Evaluation and Application of RNA-Seq by MinION. *DNA Res.* 26, 55–65. doi:10.1093/dnares/dsy038

Sessegolo, C., Cruaud, C., Da Silva, C., Cologne, A., Dubarry, M., Derrien, T., et al. (2019). Transcriptome Profiling of Mouse Samples Using Nanopore Sequencing of cDNA and RNA Molecules. *Sci. Rep.* 9, 14908. doi:10.1038/s41598-019-51470-9

Shaffer, L. (2019). Inner Workings: Portable DNA Sequencer Helps Farmers Stymie Devastating Viruses. *Proc. Natl. Acad. Sci. USA* 116, 3351–3353. doi:10.1073/pnas.1901806116

Shah, N. N., Qin, H., Yates, B., Su, L., Shalabi, H., Raffeld, M., et al. (2019). Clonal Expansion of CAR T Cells Harboring Lentivector Integration in the CBL Gene Following Anti-CD22 CAR T-Cell Therapy. *Blood Adv.* 3, 2317–2322. doi:10.1182/bloodadvances.2019000219

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., et al. (2013). Single-cell Transcriptomics Reveals Bimodality in Expression and Splicing in Immune Cells. *Nature* 498, 236–240. doi:10.1038/nature12172

Sharon, D., Tilgner, H., Grubert, F., and Snyder, M. (2013). A Single-Molecule Long-Read Survey of the Human Transcriptome. *Nat. Biotechnol.* 31, 1009–1014. doi:10.1038/nbt.2705

Sheynkman, G. M., Tuttle, K. S., Laval, F., Tseng, E., Underwood, J. G., Yu, L., et al. (2020). ORF Capture-Seq as a Versatile Method for Targeted Identification of Full-Length Isoforms. *Nat. Commun.* 11, 2326. doi:10.1038/s41467-020-16174-z

Shionyu, M., Yamaguchi, A., Shinoda, K., Takahashi, K., and Go, M. (2009). ASALPS: A Database for Analyzing the Effects of Alternative Splicing on Protein Structure, Interaction and Network in Human and Mouse. *Nucleic Acids Res.* 37, D305–D309. doi:10.1093/nar/gkn869

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., and Timp, W. (2017). Detecting DNA Cytosine Methylation Using Nanopore Sequencing. *Nat. Methods* 14, 407–410. doi:10.1038/nmeth.4184

Singh, M., Al-Eryani, G., Carswell, S., Ferguson, J. M., Blackburn, J., Barton, K., et al. (2019). High-throughput Targeted Long-Read Single Cell Sequencing Reveals the Clonal and Transcriptional Landscape of Lymphocytes. *Nat. Commun.* 10, 3120. doi:10.1038/s41467-019-11049-4

Soneson, C., Love, M. I., and Robinson, M. D. (2016). Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences. *F1000Res* 4, 1521. doi:10.12688/f1000research.7563.2

Soneson, C., Yao, Y., Bratus-Neuenschwander, A., Patrignani, A., Robinson, M. D., and Hussain, S. (2019). A Comprehensive Examination of Nanopore Native RNA Sequencing for Characterization of Complex Transcriptomes. *Nat. Commun.* 10, 3359. doi:10.1038/s41467-019-11272-z

Song, Y., Botvinnik, O. B., Lovci, M. T., Kakaradov, B., Liu, P., Xu, J. L., et al. (2017). Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Mol. Cell* 67, 148–161. doi:10.1016/j.molcel.2017.06.003

Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frébourg, T., et al. (2016). Exonic Splicing Mutations Are More Prevalent Than Currently Estimated and Can Be Predicted by Using In Silico Tools. *Plos Genet.* 12, e1005756. doi:10.1371/journal.pgen.1005756

Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., et al. (2016). Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics. *Science* 353, 78–82. doi:10.1126/science.aaf2403

Steijger, T., Abril, J. F., Abril, J. F., Engström, P. G., Kokocinski, F., Hubbard, T. J., et al. (2013). Assessment of Transcript Reconstruction Methods for RNA-Seq. *Nat. Methods* 10, 1177–1184. doi:10.1038/nmeth.2714

Strausberg, R. L., Feingold, E. A., Grouse, L., Derge, J. G., Klausner, R., Collins, F., et al. (2002). Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. USA* 99, 16899–16903. doi:10.1073/pnas.242603899

Sui, X., Kong, N., Ye, L., Han, W., Zhou, J., Zhang, Q., et al. (2014). p38 and JNK MAPK Pathways Control the Balance of Apoptosis and Autophagy in Response to Chemotherapeutic Agents. *Cancer Lett.* 344, 174–179. doi:10.1016/j.canlet.2013.11.019

Suryamohan, K., Krishnankutty, S. P., Guillory, J., Jevit, M., Schröder, M. S., Wu, M., et al. (2020). The Indian Cobra Reference Genome and Transcriptome Enables Comprehensive Identification of Venom Toxins. *Nat. Genet.* 52, 106–117. doi:10.1038/s41588-019-0559-8

Tang, A. D., Soulette, C. M., van Baren, M. J., Hart, K., Hrabeta-Robinson, E., Wu, C. J., et al. (2020). Full-length Transcript Characterization of SF3B1 Mutation in Chronic Lymphocytic Leukemia Reveals Downregulation of Retained Introns. *Nat. Commun.* 11, 1438. doi:10.1038/s41467-020-15171-6

Tapial, J., Ha, K. C. H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., et al. (2017). An Atlas of Alternative Splicing Profiles and Functional Associations Reveals New Regulatory Programs and Genes that Simultaneously Express Multiple Major Isoforms. *Genome Res.* 27, 1759–1768. doi:10.1101/gr.220962.117

Tardaguila, M., de La Fuente, L., Marti, C., Pereira, C., Pardo-Palacios, F. J., del Risco, H., et al. (2018). SQANTI: Extensive Characterization of Long-Read Transcript Sequences for Quality Control in Full-Length Transcriptome Identification and Quantification. *Genome Res.* 28, 396–411. doi:10.1101/gr.222976.117

Tian, L., Jabbari, J. S., Thijssen, R., Gouil, Q., Amarasinghe, S. L., Kariyawasam, H., et al. (2020). Comprehensive Characterization of Single Cell Full-Length Isoforms in Human and Mouse with Long-Read Sequencing. *bioRxiv*, 243543. doi:10.1101/2020.08.10.243543

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript Assembly and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation. *Nat. Biotechnol.* 28, 511–515. doi:10.1038/nbt.1621

Tress, M. L., Abascal, F., and Valencia, A. (2017). Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends Biochem. Sci.* 42, 98–110. doi:10.1016/j.tibs.2016.08.008

Treutlein, B., Gokce, O., Quake, S. R., and Südhof, T. C. (2014). Cartography of Neurexin Alternative Splicing Mapped by Single-Molecule Long-Read mRNA Sequencing. *Proc. Natl. Acad. Sci. USA* 111, E1291–E1299. doi:10.1073/pnas.1403244111

Uapinyoying, P., Goecks, J., Knoblach, S. M., Panchapakesan, K., Bonnemann, C. G., Partridge, T. A., et al. (2020). A Long-Read RNA-Seq Approach to Identify Novel Transcripts of Very Large Genes. *Genome Res.* 30, 885–897. doi:10.1101/gr.259903.119

Ule, J., and Blencowe, B. J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol. Cell* 76, 329–345. doi:10.1016/j.molcel.2019.09.017

Ullrich, B., Ushkaryov, Y. A., Südhof, T. C., Beate Ullrich, A., Ushkaryov, Y., and Südhof, T. (1995). Cartography of Neurexins: More Than 1000 Isoforms Generated by Alternative Splicing and Expressed in Distinct Subsets of Neurons. *Neuron* 14, 497–507. doi:10.1016/0896-6273(95)90306-2

Vitting-Seerup, K., and Sandelin, A. (2017). The Landscape of Isoform Switches in Human Cancers. *Mol. Cancer Res.* 15, 1206–1220. doi:10.1158/1541-7786.mcr-16-0459

Volden, R., and Vollmers, C. (2020). *Highly Multiplexed Single-Cell Full-Length cDNA Sequencing of Human Immune Cells with 10X Genomics and R2C2*. bioRxiv, 902361. doi:10.1101/2020.01.10.902361

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* 456, 470–476. doi:10.1038/nature07509

Weatheritt, R. J., Sterne-Weiler, T., and Blencowe, B. J. (2016). The Ribosome-Engaged Landscape of Alternative Splicing. *Nat. Struct. Mol. Biol.* 23, 1117–1123. doi:10.1038/nsmb.3317

Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., et al. (2017). Comprehensive Comparison of Pacific Biosciences and Oxford Nanopore Technologies and Their Applications to Transcriptome Analysis. *F1000Res* 6, 100. doi:10.12688/f1000research.10571.2

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., et al. (2019). Accurate Circular Consensus Long-Read Sequencing Improves Variant Detection and Assembly of a Human Genome. *Nat. Biotechnol.* 37, 1155–1162. doi:10.1038/s41587-019-0217-9

Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., et al. (2019). Nanopore Native RNA Sequencing of a Human Poly(A) Transcriptome. *Nat. Methods* 16, 1297–1305. doi:10.1038/s41592-019-0617-2

Wyman, D., Balderrama-Gutierrez, G., Reese, F., Jiang, S., Rahmanian, S., Forner, S., et al. (2019). *A Technology-Agnostic Long-Read Analysis Pipeline for Transcriptome Discovery and Quantification*. bioRxiv, 672931. doi:10.1101/672931

Yap, K., and Makeyev, E. V. (2016). Functional Impact of Splice Isoform Diversity in Individual Cells. *Biochem. Soc. Trans.* 44, 1079–1085. doi:10.1042/bst20160103

You, R., Zhang, Z., Xiong, Y., Sun, F., Mamitsuka, H., and Zhu, S. (2018). GOLabeler: Improving Sequence-Based Large-Scale Protein Function Prediction by Learning to Rank. *Bioinformatics* 34, 2465–2473. doi:10.1093/bioinformatics/bty130

Zhang, C., Zhang, B., Lin, L.-L., and Zhao, S. (2017). Evaluation and Comparison of Computational Tools for RNA-Seq Isoform Quantification. *BMC Genomics* 18, 583. doi:10.1186/s12864-017-4002-1

Conflict of Interest: RP, JG and MC have received support from ONT to present their findings at scientific conferences. ONT played no role in study design, execution, or publication.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

GLOSSARY

5mC 5-methylcytosine
AS alternative splicing
BCR B-cell receptor
CLL chronic lymphocytic leukemia
CCS circular consensus sequencing
CDS coding sequence
DIE differential isoform expression
DIU differential isoform usage
GWAS genome-wide association study
GTEX genotype-tissue expression
iPSC induced pluripotent stem cell
lncRNAs long-noncoding RNAs
LCS long-read RNA CaptureSeq
LRS long-read sequencing
LR scRNA-Seq long-read single cell sequencing
m6A N6-methyladenosine
NMD nonsense-mediated decay
OCS ORF Capture-Seq
ONT Oxford Nanopore Technology
PacBio Pacific Biosciences
R2C2 Rolling Circle Amplification to Concatemeric Consensus
SiT Spatial Isoform Transcriptomics
SRS short-read sequencing
SR scRNA-Seq short-read single cell sequencing
SMRT single molecule, real-time
scRNA-Seq single cell sequencing
sgRNA subgenomic RNA
TCR T-cell receptor
TSS transcription start site
UMI unique molecular identifier
UCSC University of California Santa Cruz
UTR untranslated region
VZV varicella-zoster virus
VastDB vertebrate alternative splicing and transcription database