

Bioinformatics frameworks for single-cell long-read sequencing: unlocking isoform-level resolution

Saloni Bhatia^{1,2}, Matt A. Field ^{2,3,4}, Lionel Hebbard ^{2,5,6}, Ulf Schmitz ^{1,2,7,*}

¹Computational Biomedicine Lab, College of Science and Engineering, James Cook University, 1 James Cook Drive, Townsville, QLD 4811, Australia

²Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, 14-88 McGregor Road, Smithfield, QLD 4878, Australia

³Immunogenomics Lab, Garvan Institute of Medical Research, 384 Victoria St, Darlinghurst, NSW 2010, Australia

⁴Menzies School of Health Research, Charles Darwin University, Red 9, Casuarina campus, Univ Drive North, Casuarina, NT 0811, Australia

⁵Storr Liver Centre, Westmead Institute for Medical Research, Westmead Hospital and University of Sydney, 176 Hawkesbury Rd, Westmead, Westmead, NSW 2145, Australia

⁶Australian Institute for Tropical Health and Medicine, 1 James Cook Drive, Townsville, QLD 4811, Australia

⁷Centenary Institute, The University of Sydney, Building 93, Royal Prince Alfred Hospital Missenden Rd, Camperdown, NSW 2006, Australia

*Corresponding author. Computational Biomedicine Lab, College of Science and Engineering, James Cook University, 1 James Cook Drive, Townsville, Queensland, Australia. E-mail: ulf.schmitz@jcu.edu.au

Abstract

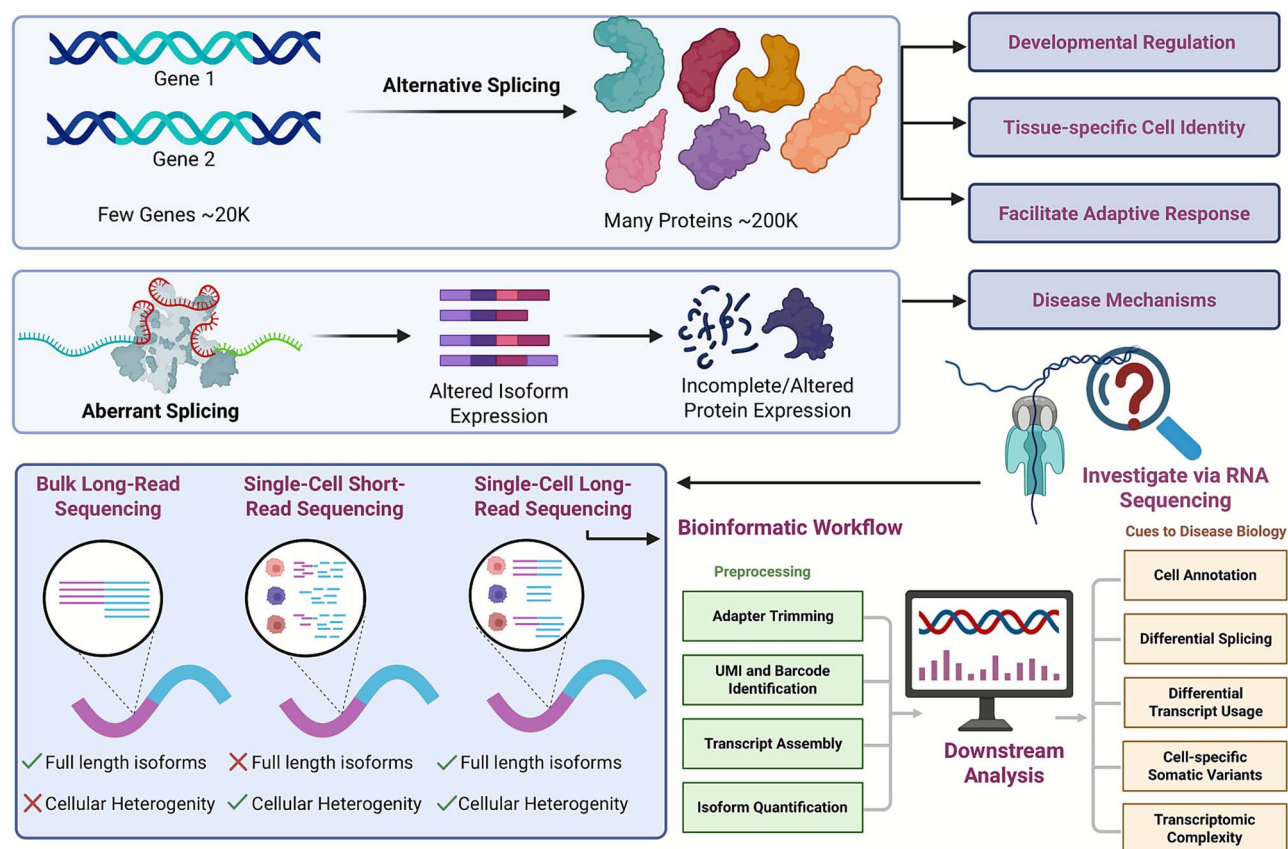
Alternative splicing (AS) plays a key role in regulating gene expression, and its dysregulation is implicated in numerous human diseases, including cancer. While bulk RNA sequencing has advanced our understanding of AS, it cannot capture cellular heterogeneity or reliably reconstruct full-length isoforms, both of which underpin disease mechanisms and therapeutic responses. Single-cell RNA sequencing (scRNA-seq) is an established and a powerful approach to examine AS landscapes at single-cell resolution, enabling the identification of cell-specific aberrant splicing events that may contribute to disease. However, conventional scRNA-seq is limited by short read lengths, often preventing an accurate reconstruction of full-length transcript isoforms. This limitation is addressed by long-read RNA-seq (lrRNA-seq), which can sequence full-length RNA molecules, some exceeding 100 000 nucleotides in length. Thereby, lrRNA-seq enables more accurate characterization of isoform diversity, identification of novel splice variants, quantification of percent spliced-in values, and detection of fusion transcripts. The convergence of single-cell resolution and third-generation sequencing technologies has led to the development of single-cell long-read sequencing (SCLR-seq), a powerful approach that addresses the key constraints of bulk short-read RNA-Seq by providing isoform-level resolution and cell-type specificity. This review explores the growing utility of SCLR-seq, highlighting recent developments in bioinformatics tools and pipelines designed for SCLR-seq data analysis. We discuss how this emerging technology is transforming our understanding of isoform regulation and aberrant splicing in human diseases, and its potential to uncover novel diagnostic and therapeutic targets.

Received: August 15, 2025. Revised: October 24, 2025. Accepted: November 14, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Graphical Abstract



Keywords: alternative splicing; single-cell long-read sequencing; isoform quantification; differential isoform expression

Introduction

Important biological processes such as cell development and differentiation are directed by gene expression networks, often modulated by alternative splicing (AS) with over 30% of tissue-dependent transcript variants constituted by local splicing events [1]. In this context, alternative isoforms can be generated from the same gene by either differential processing of the pre-mRNA or by using alternative transcription start/end sites. A set of carefully balanced regulatory mechanisms interact with each other to tightly regulate AS spatiotemporally [2], meaning that any disruption in AS can lead to or mediate diseases. Numerous human hereditary diseases and cancers are associated with aberrant AS [3, 4].

Profiling AS and isoform expression via RNA sequencing

To understand the role of AS in human diseases, it is important to accurately map and quantify AS events and decipher their functional significance. High-throughput RNA-Seq of patient-derived tissue samples has been useful for generating gene expression profiles that could be mined, e.g. for diagnostic and prognostic markers for clinical management. Such transcriptome profiling studies have largely relied on bulk short-read RNA sequencing (RNA-Seq) [5]; however, this approach fails to capture the genomic heterogeneity between different cell types within the same tissue.

In fact, cellular heterogeneity has a significant impact on clinical outcomes altering the efficacy of some therapies [6]. To capture this heterogeneity, single-cell RNA-seq (scRNA-seq) is used to examine transcriptomic cell-to-cell variation. While bulk RNA-seq data are dominated by abundant cell types relative to rare cell types, scRNA-seq can capture information from both abundant and relatively rare cell types.

Beyond cellular heterogeneity and rare cell types, single-cell transcriptomics provides insights into cellular architecture and dynamic transcriptomic changes across time or experimental conditions. Every tissue contains functionally and morphologically diverse populations of cells in different states and spatial locations. Cell architecture in disease states has been investigated very well in tumors, e.g. many biological processes including development, immune responses, and cancer evolution are driven by cell-specific transcriptomic programs. Single-cell resolution is therefore critical to capture this diversity. For instance, rare cell types may express unique isoforms that are undetectable in bulk, and clonal populations within tumors may show distinct splicing patterns or mutational profiles.

However, transcript-level analysis in droplet-based scRNA-seq platforms such as 10× Genomics is inherently limited by protocol-induced biases. Depending on the chemistry employed (3' or 5' capture), these protocols selectively enrich for one end of the transcript, resulting in biased coverage concentrated at the 3' or 5' end and incomplete representation of the full transcript body.

This results in uneven coverage across transcripts, often missing critical splice junctions or transcript start/end sites required for accurate isoform resolution.

Additionally, one end of each read is typically reserved for sequencing the cell barcode (CB) and unique molecular identifier (UMI), leaving only the remaining portion of the read, often a single-end read of ~90–100 bp, for transcript alignment. This reduced effective read length limits the ability to uniquely map reads, especially in repetitive or homologous regions, thereby inflating multi-mapping issues. Moreover, reads originating from different isoforms or spanning complex homologous regions cannot be reliably distinguished, hindering accurate identification of novel isoforms and full-length gene fusion events [7]. In contrast to bulk short-read RNA-Seq, where paired-end reads and higher RNA input provide uniform transcript coverage and robust isoform reconstruction, single-cell protocols are impacted by sparse coverage and high dropout rates, leading to lower sensitivity and reduced recovery of transcript diversity. However, scRNA-seq studies have revealed that AS exhibits a high level of heterogeneity across cell types [8–10], and high cell-type specificity [11], particularly in cancer [12–14].

Long-read RNA-Seq

Long-read RNA-seq (lrRNA-seq) captures full-length transcripts generating reads of up to hundreds of thousands of nucleotides in length. Consequently, lrRNA-seq has revealed many structurally diverse transcripts and novel isoforms [15]. Because long reads can span entire transcript molecules, lrRNA-seq enables the direct observation of co-occurring events within single isoforms that are often fragmented or entirely missed in short-read data. This includes the coordination between AS events and other transcript elements such as transcription start sites (TSSs) and poly(A) sites. For example, long-read studies have shown that alternative promoter usage can be tightly linked to downstream splicing patterns, with initiation at specific TSSs predisposing transcripts to exon or poly(A) site choices [16, 17].

A combination of reduced error rates, lower costs, and increased throughput has driven greater adoption of lrRNA-seq, also widely referred to as third-generation sequencing (TGS), for bulk-based and single-cell sequencing [15, 18]. Single-cell long-read RNA-seq (SCLR-seq) provides an opportunity to capture full-length isoforms at single-cell level. Unlike short-read single-cell approaches, which rely on incomplete transcript capture and isoform inference, SCLR-seq gives a comprehensive and annotation-independent view of the transcriptome, revealing the full extent of isoform diversity, AS patterns, and transcript usage in cell.

The complexity of isoform expression further fine-tunes cell identity as it contributes to the diversity of phenotypes across cells, cell types, tissues and organs among individuals [19]. With SCLR-seq, we are moving beyond differential gene expression as a source of cell diversity to differential isoform expression as differences between cell populations. These isoform level signatures allow for finer discrimination between cell subtypes, transitional states, or disease-associated phenotypes. SCLR-seq enables more detailed characterization of regulatory mechanisms such as allele-specific isoform expression, coordinated splicing, and other RNA processing events. It can also help resolve ambiguity in complex genomic loci, such as those involving bidirectional promoters or overlapping transcripts including long noncoding, which are often difficult to distinguish using short-read data.

Together, these capabilities enable SCLR-seq to provide a far more detailed and functionally relevant view of the transcriptome

than previously possible. This review discusses the applications and methods of SCLR-seq with a focus on bioinformatics data analysis pipelines.

The importance of AS in human health

Nearly all protein coding genes in the human genome undergo AS [20]. AS events facilitate the expression of transcript isoforms in a developmental, cell-specific and tissue-specific manner [21], with ~95% of multi-exon genes undergoing splicing to produce an average of 14.6 isoforms per gene [22, 23]. Different mRNA isoforms encode proteins or non-coding RNAs that may differ in their function, structure, and localization with isoforms even having opposing functions in some cases. For example, most apoptosis-related genes, such as *BCL2L1* (*BCL2* Like 1) and *CASP2* (caspase-2) have multiple isoforms, some functioning to promote apoptosis while others are anti-apoptotic [24, 25], with the ratio of isoforms expressed deciding the fate of a cell. AS also plays a regulatory role in gene expression by influencing mRNA stability. It can affect the half-life of mRNA molecules by retaining or omitting certain regulatory elements, or by including a premature stop codon, causing early degradation of transcripts [26, 27]. Splicing also affects translation rates by influencing the processing and stability of mRNA, e.g. through the recruitment of exon junction complexes, which help engage translation initiation factors and enhance ribosome loading [28].

AS is closely related to human health and often caused by genetic variants (Fig. 1). Single nucleotide variants (SNVs) such as missense mutations, stop-gain mutations, small insertions or deletions can disrupt normal splicing and cause disease [29]. These variants may interfere with splicing signals, enhancers or silencers within the pre-mRNA, leading to the production of aberrant mRNA transcripts and proteins. Such *cis*-acting variants are estimated to constitute 15%–30% of disease-causing mutations in human [30]. In contrast, SNVs that effect *trans*-regulatory elements can induce widespread splicing changes by altering splicing of many target genes; however, they are rare compared to variants in *cis* elements. Deep intronic variants can create cryptic splice sites leading to abnormal transcript isoforms [31]. Additionally, structural variants like translocations, inversions, copy number variants, and gene fusions can alter splicing by rearranging the order of exons and introns, introducing cryptic splice sites, deleting exons, and disrupting splice regulatory mechanisms [29].

Variants in consensus splice site (GU at 5' end of the intron and AG at its 3' end) as well as in *cis*- or *trans*- acting elements can cause or contribute to many human diseases and are collectively referred to as spliceopathies [29, 32]. For example, germline variants in *trans*-acting elements of the spliceosome were recently found to lead to converging neurodevelopmental phenotypes of developmental delay, intellectual disability, and autism [33]. Other diseases associated with aberrant splicing include bipolar disorder, schizophrenia [34], cancer [35–37], amyotrophic lateral sclerosis [38], diabetes [39], and myotonic dystrophy [40]. Splicing variants can also impact translation rates by altering binding sites for cap-binding proteins, poly(A)-binding proteins, RNA-binding proteins, or sequences upstream of the initiator codon. For example, a splicing-associated A-to-G substitution at the +1 position converted an upstream adenine-uracil-guanine (AUG) codon to a guanine-uracil-guanine (GUG) in the upstream open reading frame (uORF) of *EPHB1*, which reduced uORF translation and increased translation at the main ORF [41, 42].

Aberhart isoform expression can cause development-related pathologies such as Stickler syndrome [43], accelerate aging [44],

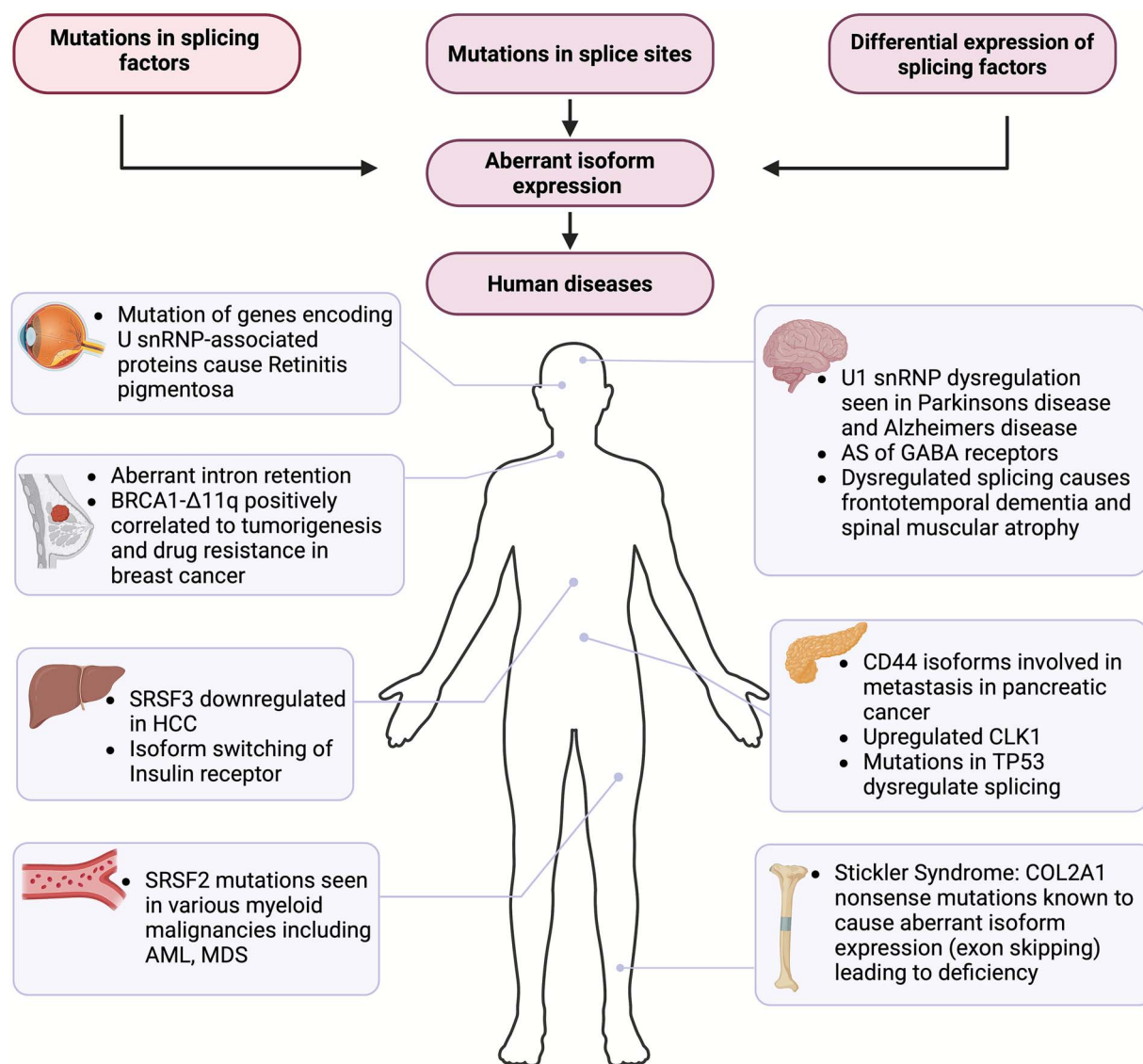


Figure 1. Examples of human diseases caused due to dysregulated alternative splicing. Created with BioRender.com.

impair functioning of the heart and brain [45]. It is also linked to various cancers and their hallmarks, including uncontrolled proliferation, metastasis, immune invasion, and oncological transformation [35, 46, 47]. For example, splice variants in CD44, a transmembrane glycoprotein expressed on the surface of many cells, are known to be involved in tumor metastasis in pancreatic cancer and breast cancer [48]. Missense mutations in TP53 (tumor suppressor 53) have been associated with the dysregulation of AS in pancreatic cancer [49]. It has also been noted that the expression of factors involved in pre-mRNA splicing is most altered in cancer [50]. By example, SRSF3 is a splicing factor (SF) responsible for AS of target genes involved in glucose and lipid metabolism, in addition to roles in cytoskeleton organization and cellular growth. It is downregulated in hepatocellular carcinoma (HCC), and its knockdown in mouse hepatocytes promotes the development of metastatic HCC [51]. SRSF2 is an RBP that plays important roles in splicing of mRNA precursors. SRSF2 mutations are the most common splicing machinery mutations identified in various myeloid malignancies like acute myeloid leukaemia, myelodysplastic syndrome, chronic myelomonocytic leukaemia, and myeloproliferative neoplasms [52].

Other sources of isoform diversity

In eukaryotic cells, isoform diversity is substantially expanded by context-dependent alternative RNA processing events. These include the differential use of TSSs, splice sites, and polyadenylation sites [53]. Among these, alternative polyadenylation plays a critical role in shaping transcript isoforms by altering 3' untranslated region length or generating distinct protein-coding variants thereby influencing mRNA stability, localization, and translational efficiency. More than half of human genes are estimated to undergo alternative 3'-end processing [54].

In addition to canonical processing, transcriptomic complexity is further enhanced by posttranscriptional RNA modifications. The most prevalent of these is N6-methyladenosine (m6A), the methylation of adenosine residues. m6A modifications impact multiple aspects of RNA metabolism including splicing, nuclear export, stability, and translation thereby adding an additional regulatory layer that contributes to isoform and functional diversity [55].

Further complexity arises from RNA editing, particularly adenosine-to-inosine (A-to-I) editing mediated by Adenosine Deaminase Acting on RNA (ADAR) enzymes, which can recode

transcripts, alter splice sites, or modulate RNA structure [56]. Transcriptional readthrough and the formation of chimeric transcripts often spanning adjacent genes also expand the isoform landscape, especially in certain disease contexts, such as cancer [57]. Complex gene architectures including overlapping or antisense transcripts contribute to transcriptomic ambiguity that is difficult to resolve with short-read methods but can be clarified using long-read sequencing [58, 59].

TGS platforms for SCLR-seq

Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) are the two prominent platforms for TGS that are increasingly adopted in SCLR-seq experiments. PacBio sequencing provides high accuracy, but its substantial setup and sequencing costs restrict its use. In contrast, ONT sequencers are affordable, provide high throughput, but come with a higher sequencing error rates impacting base call accuracy [60].

PacBio's single-molecule real-time (SMRT) sequencers (Sequel, Sequel II, Sequel IIe, Revio, and Vega) detect nucleotides as they are incorporated in real-time by monitoring fluorescence pulses emitted as DNA polymerase synthesizes a complementary strand within nanophotonic zero-mode waveguides [61]. Sequel II can produce up to 24 Gb of data per day, whereas the Revio system can produce up to 360 Gb data per run [62]. Both systems support circular consensus sequencing, in which a single DNA molecule is repeatedly sequenced in a circularized format to generate a highly accurate consensus read. PacBio claims that using this method leads to read accuracies of 99.9% and 99.95% in their Sequel II and Revio sequencers, respectively. However, the Sequel II and Sequel IIe systems are no longer available for sale.

When using 8M SMRT Cells (containing 8 million zero-mode waveguides), up to 2 million HiFi reads can be obtained. The MAS-Iso-Seq approach developed by Al'Khafaji *et al.* reaches nearly 40 million HiFi reads per run on the Sequel IIe sequencer [63]. PacBio adopted MAS-Iso-Seq in their Kinnex library preparation kits, which produce sequencing ready SMRTbell libraries for use on either Revio or Vega systems, which are their latest models in PacBio's long-read sequencer fleet. Revio can generate up to 80–100 million reads per run and has 20× the compute power of the Sequel IIe system, whereas Vega is a benchtop sequencer that can generate 50–60 million reads per run.

ONT's sequencing device range includes the portable MinION, and the more stationary GridION, and PromethION, which can be used with one or more different types of flow cells including MinION (2048 pores), Flongle (512 pores), and PromethION (10 240 pores). A single PromethION flow cell can generate up to 250 million reads theoretically, making it a powerful tool for high-throughput applications [64]. However, such yields are not typical in practice due to factors such as library quality, sample complexity, and flow cell performance. Nanopore sequencing used to encounter high error rates; however, with recent advancements in flow cell chemistry and improvements in basecalling algorithms, ONT reported accuracies of up to 99% [65].

SCLR-seq library preparation methods

Like short-read scRNA-seq, libraries for SCLR-seq can be prepared using droplet based, plate based, or microwell-based methods. Multiple 10× Genomics kits have demonstrated compatibility with both PacBio [66, 67] and ONT [68, 69] protocols (Table 1). The 10× Genomics Chromium platform for single-cell analysis can facilitate capture of 10 000 cells per library and enable rare cell

type discovery; however, it suffers from a high dropout rate, especially for genes with low expression [70]. Two recently introduced protocols for SCLR-seq, SciSOr-Seq [67] and ScNaUmi-seq [69], adopt 10× Genomics' Chromium kits to barcode cells and to split them for complementary long- and short-read sequencing. Both protocols use the short reads for error correction and guiding barcode assignment of the long reads. Similarly, the FLT-Seq library preparation method uses a modified Chromium 10× droplet-based workflow in which a subset of gel beads-in-emulsions (GEMs) (e.g. 10%–20%) is diverted for long-read sequencing, while the remainder proceeds with standard short-read sequencing. This “hybrid subsampling” strategy allows leveraging the high throughput and low cost of short-read sequencing for cell clustering and barcode assignment, while still gaining full-length isoform data for a subset of cells [71]. The RAGE-seq [72] protocol, introduced by Singh *et al.* in 2019, in addition to sequencing both short and long reads, performs a targeted capture prior to sequencing to improve depth of coverage. However, a distorted distribution of CBs and UMIs was observed due to polymerase chain reaction (PCR) cycles and error rates of ONT. Rapid Capture Hybridization Sequencing is a targeted capture hybridization method applied to (single-cell) full-length cDNA, where probes are used to enrich transcripts of interest, which are then sequenced with long reads [73]. Long-read targeted scRNA-seq is a more recent technique to combine targeted long-read sequencing of selected transcripts with the retention of single-CBs/UMIs, using a custom probe panel, thereby focusing sequencing resources on molecules of interest (e.g. low-abundance or functionally important isoforms) [74]. Single cells isolated by Drop-seq are sequenced on an ONT platform in a protocol termed scCOLOR-seq [64], which aims at reducing errors in barcodes. This protocol incorporates bi-nucleotide repeats in barcode sequences which enables detection and correction of sequencing errors in barcodes. It demonstrated significant improvement in barcode accuracy across different cell types, yielding reliable isoform detection, proving to be a robust method to perform quantitative long-read transcript sequencing on large numbers of cells. Nonetheless, it requires customized synthesis of gel-beads and prior knowledge of true barcodes [64].

Smart-seq2 [75] is a plate-based method that was initially developed for full-transcript coverage on Illumina next-generation sequencing platforms. Different research groups have built on the Smart-seq2 protocol to enable efficient sequencing on TGS platforms. The R2C2 (Rolling Circle to Concatemeric Consensus) protocol uses the Tn5Prime method to prepare libraries, which is a modification of the smart-seq2 method that uses a distinct template switch oligo (TSO) containing 7-nt sample indexes during reverse transcription [76]. By using the circular consensus principle, the protocol achieves ~96% sequencing accuracy with ONT-based long-read sequencing. Plate-based methods limit the number of cells that can be sequenced: R2C2 sequenced 1–96 cells, while SCAN-seq2 [77], another plate-based method compatible with ONT sequenced 96–960 cells. However, it is possible to capture more cells by processing more plates in parallel, although this increases cost and handling complexity.

Microwell-based methods such as Parse Biosciences' Evercode™ have also been used with PacBio [78] and ONT [79] (Table 1). Microwell-based methods use combinatorial indexing and alleviate the use of microfluidic devices. They have higher throughput, but lengthy hands-on time, which could lead to inconsistencies stemming from human error [80].

Table 1. Available single-cell kits and their compatibility with ONT and PacBio TGS platforms

Available kits	Company	Compatibility with ONT	Compatibility with PacBio
Chromium Next GEM Single Cell 3' Kit v3.1	10× Genomics	Validated and supported	Validated and supported
Chromium Next GEM Single Cell 5' Kit v2	10× Genomics	Validated and supported	Validated and supported
Chromium Next GEM Single Cell 5' Kit v2—V(D)J library	10× Genomics	Currently being tested	Incompatible
Chromium GEM-X Single Cell 5' Kit v3	10× Genomics	Not tested	Compatible with Kinnex kit, not tested for MAS-Seq kit
Chromium GEM-X Single Cell 3' Kit v4	10× Genomics	Not tested	Not tested
Chromium Next GEM Single Cell Multiome ATAC + Gene Expression	10× Genomics	Currently being tested	Compatible with both kits
Chromium Next GEM Single Cell ATAC Kit v2	10× Genomics	Not tested	Incompatible
Chromium Next GEM Single Cell Fixed RNA Sample Preparation Kit	10× Genomics	Incompatible	Incompatible
Evercode Whole Transcriptome v3	Parse Biosciences	Currently being tested	Compatible, tested in-house
ScaleBio Single Cell RNA Sequencing Kit v1.1	Scale Biosciences	Not tested	Not tested/not compatible (not specified)
ScaleBio™ Single Cell Methylation Kit	Scale Biosciences	Not tested	Not tested/not compatible (not specified)
Illumina Single Cell 3' RNA Prep	Fluent BioSciences (Illumina)	Not tested	Tested with MAS-seq

Bioinformatics analysis of SCLR-seq data

Analysis of SCLR-seq data involves a series of computational steps, each designed to interpret the transcriptome at single-cell resolution using full-length reads. These typically include: (i) extraction of CBs and UMIs, (ii) read alignment and transcriptome assembly, (iii) quantification of genes and isoforms, (iv) cell-type identification and clustering, and (v) downstream analyses such as isoform classification, differential isoform usage, and differential splicing (Fig. 2a). Downstream analyses like AS analysis and isoform classification can be performed using existing tools for single-cell [81] or long-read data [82]; however, functionality might be limited when applied to SCLR-seq data. While many of these steps are well-established for bulk or short-read scRNA-seq, SCLR-seq introduces unique challenges. These include higher per-read error rates, barcode or UMI loss, lack of a defined read structure (e.g. 3' bias), and difficulties in assigning transcript boundaries due to read truncation or chimera formation. As a result, conventional single-cell or long-read tools often require adaptation or entirely new methods for accurate SCLR-seq analysis.

A growing number of tools and pipelines have been developed to support SCLR-seq analysis. These range from end-to-end workflows, which automate most of the pipeline using tools focused on individual tasks such as demultiplexing, alignment, or isoform quantification. The sections below follow the typical analysis workflow and, for each step, outline the key conceptual goals, challenges unique to SCLR-seq, and tools currently available. Both PacBio and ONT also provide their own workflows to perform end-to-end analysis. This section will discuss current trends in methods used to interpret SCLR-seq data.

Processing of raw sequencing signals

Raw sequencing data generated from ONT platforms are stored in POD5, FAST5, or SLOW5 [83] format, which contains the squiggle signal produced via disruption in ionic current flow as cDNA fragments traverse nanopores embedded in a membrane. The raw signal is then analyzed by a basecaller software that uses deep learning models to determine the nucleotide sequence encoded in each read. ONT, developed its own high-accuracy basecaller

dorado (github.com/nanoporetech/dorado), which stores sequencing reads in FASTQ files and has built-in functions for demultiplexing, as well as adapter- and poly-A trimming; however, it cannot be used to extract CBs and UMIs.

PacBio HiFi reads, which include fluorescent pulses, need to be divided into separate segmented reads (S-reads) that correspond to the original cDNA sequences. This can be done on PacBio's analysis software SMRT Link (www.pacb.com/smrt-link/). Read segmentation includes converting fluorescent pulses into continuous long reads, then into subreads which are stored as unaligned BAM files. This is finally converted to a continuous circular read, which is compatible for use in PacBio's downstream Iso-Seq workflow.

CB and UMI extraction from reads

Accurate identification of reads in scRNA-seq data relies on accurate recovery of CBs and UMIs, which are incorporated during library preparation. In droplet-based methods like 10× Genomics, each cDNA molecule is tagged with a CB to indicate its cell of origin and a UMI to uniquely label that original molecule and help identify and collapse PCR duplicates (Fig. 2b and c).

In short-read platforms, CB and UMI sequences are predefined, and a whitelist of known barcode sequences is provided, enabling straightforward extraction and error correction. Barcode recovery becomes challenging when sequenced in long-read platforms with several complications: (i) high error rates in long-read platforms complicate sequence matching and increase false positives. (ii) Barcodes may be partially or incorrectly sequenced, leading to loss of whitelist matches. (iii) Chimeric reads. Chimeric cDNA reads can be identified if they contain: (a) two adjacent internal poly-A signals with more than 200 nucleotides between them and are flanked by an adapter, (b) one internal adapter with an adjacent template switch oligo (TSO), or (c) two adjacent TSOs. TSOs are adapters used to generate the template strand from the cDNA (Fig. 2b).

Inaccurate detection of barcodes and UMIs has a direct impact on isoform discovery and quantification. It can cause false detection of spurious new isoforms and inaccurate quantification. To prevent these errors, most studies sequence the same cDNA library with both short-read and long-read methods. Short reads

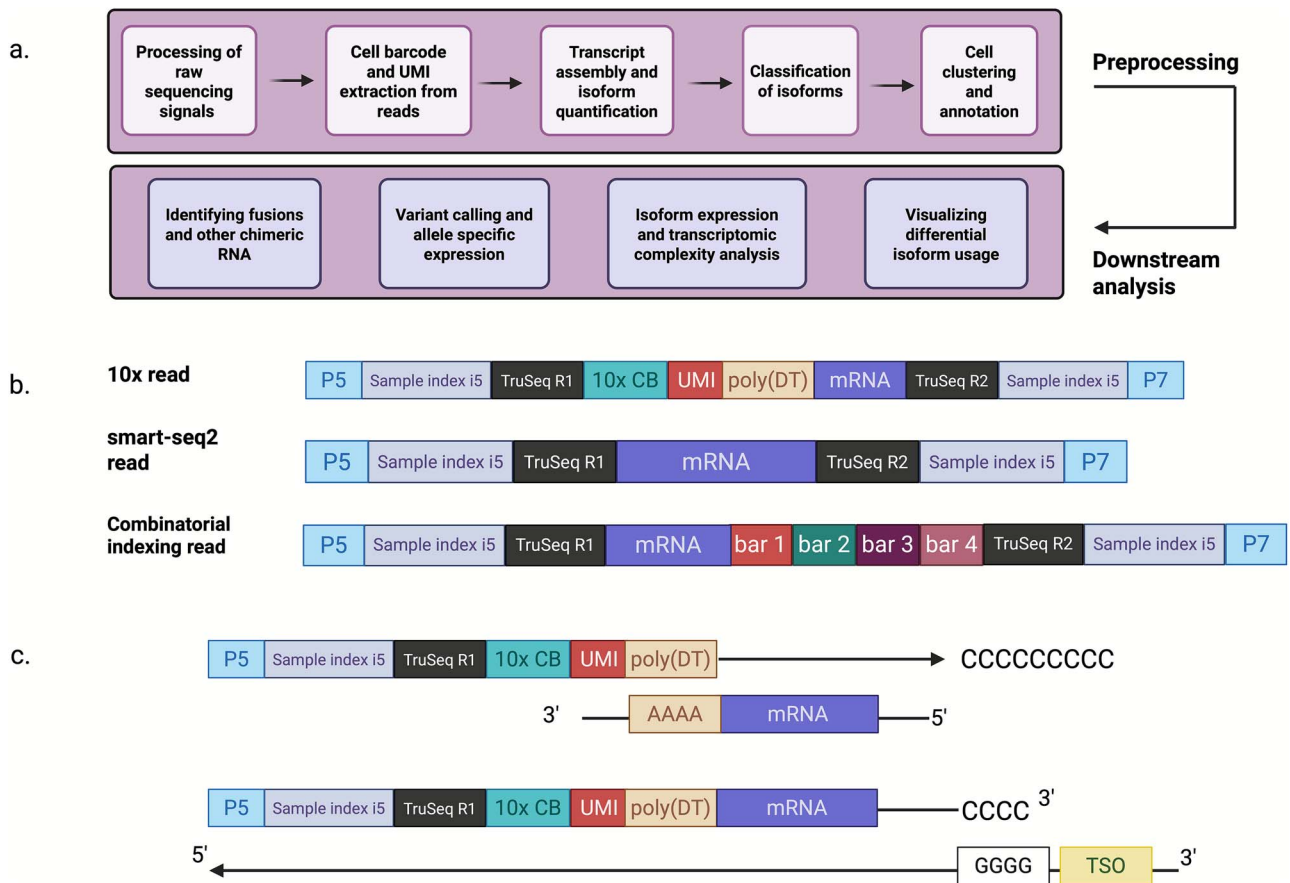


Figure 2. SCLR-seq data handling. (a) Basic steps involved in data analysis of SCLR-seq data. For a more exhaustive list see [Table 2](#) and [long-read-tools.org](#) [82]. (b) Sequencing read structures of different single-cell library preparation methods. Most droplet-based methods such as 10× Genomics utilize a CB and a UMI. Methods that use combinatorial indexing such as Parse Biosciences incorporate four barcodes for a read. The combination of these barcodes helps differentiate between cells and genes. (c) The architecture of single-cell long-reads. In the 3' kit of 10×, cDNA is sequenced using the poly-A tail; thus, it is sequenced from the 3' to 5' direction. The parent strand is generated by using a template switch oligo (TSO). The TSO causes the reverse transcriptase enzyme to switch from transcribing the RNA template sequence to transcribing the TSO sequence, generating the complementary cDNA strand. Created with [BioRender.com](#).

are used to construct reference barcodes and aid in barcode recovery from long reads. Tools like *SiCeLoRe* [69] and *scTagger* [84] allow demultiplexing of single-cell long reads from accompanying short reads. However, SCLR-seq-specific tools now facilitate barcode detection without the need for accompanying short reads.

Several tools have been developed for CB and UMI detection. ONT's *epi2me-labs/wf-single-cell* pipeline (github.com/epi2me-labs/wf-single-cell) supports barcode extraction, UMI parsing, and deduplication, using expected sequence motifs like poly-A tails and 10× adapters to orient reads. Similarly, *BLAZE* [85] identifies the poly-T signal and 10× adapter to extract CB and UMI tags directly from raw reads, saving them in a format compatible with downstream tools like *FLAMES* [71]. *BLAZE* does not require a known whitelist and can tolerate sequencing noise by applying a configurable edit distance during correction. *SiCeLoRe* further addresses the issue of chimeras by scanning for repeated adapter or poly-A motifs within long reads and splitting them into putative single-transcript molecules. This is particularly important in full-length sequencing protocols that may capture multiple cDNAs within a single read. *Flexiplex* [86], like *BLAZE*, works without known barcodes and also splits chimeric reads like *SiCeLoRe*. All these tools work with UMI-based single-cell protocols. PacBio's *Iso-Seq* workflow [87] uses primers and the poly-A tail to orient the read into a 5'–3' direction, prior to removing them using

lima. CBs and UMIs are then extracted and corrected using a reference list, followed by read deduplication based on corrected values. Microwell-based protocols like Parse Biosciences [88] that use combinatorial indexing have their own custom pipeline for barcode handling. No independent tool is available yet to facilitate this. A comparison of tools for barcode and UMI extraction, correction, and chimera handling in SCLR-seq is shown in [Table 3](#).

UMI error correction strategies include clustering UMIs within a fixed edit distance. Edit distance is defined as the number of nucleotide changes (insertions, deletions, or substitutions) required to match two sequences. An appropriate threshold depending on single-cell protocol/sequencing platform is used to offer a good balance between efficiency and specificity to allow minor corrections accounting for sequencing errors. Read inflation can also occur when the number of UMIs is overestimated in cases of high gene expression and high PCR amplification. *Longcell* [89] utilizes PCR duplicates to correct truncation and mapping errors as a UMI recovery method and applies an iterative clustering procedure to cluster putative UMI sequences. A recent study used clustering-based methods (e.g. *starcode*) to group similar barcodes and correct them to consensus sequences, while UMIs were corrected using graph-based models that incorporated cDNA length and sequence similarity to collapse amplification artifacts [90].

Transcript assembly and isoform quantification

Transcript assembly and isoform quantification are central to understanding transcriptomic diversity at single-cell resolution. The aim of this step is to reconstruct full-length transcripts or isoforms from mapped reads and assign expression levels to each isoform in individual cells. While known challenges associated with both single-cell and long-read sequencing persist, including data sparsity, high dropout rates, sequencing errors and read length bias, SCLR-seq introduces additional complexities: (i) sparse coverage due to insufficient reads per cell makes transcript reconstruction more uncertain. (ii) The accurate retention and handling of CBs and UMIs throughout the analysis are critical, as any loss or misassignment can lead to incorrect cell and transcript quantification. (iii) Isoform ambiguity arising from clustering approaches used or from sequence features, such as homology, complicates accurate isoform quantification.

For alignment, the long-read mapper *Minimap2* is widely used [91]. It supports spliced alignment and retains CB and UMI (UB) tags in the BAM output files, making them compatible with downstream isoform assembly tools. While tools such as *Bambu*, *StringTie2*, *TALON*, *FLAIR*, and *IsoQuant* have been extensively benchmarked for isoform discovery and quantification in bulk long-read datasets, they are not inherently designed for SCLR-seq, as they cannot track reads back to individual cells without additional preprocessing [92–95]. Even with barcode-aware preprocessing steps, these tools remain UMI-unaware, meaning they cannot perform accurate molecular deduplication at the isoform level. scRNA-seq tools that are UMI-aware such as *UMI-tools* or *Alevin* are built for short-read data and do not support full-length isoform reconstruction [96, 97]. The need for tools and pipelines that simultaneously support long-read isoform discovery and UMI-aware quantification highlights a major computational challenge in SCLR-seq analysis.

There is a growing need for isoform-aware UMI deduplication methods as well. Isoform discovery algorithms usually cluster similar isoforms into groups and collapse each cluster into a single representative isoform. While effective for reducing noise, this approach risks collapsing distinct isoforms into a single dominant isoform, leading to inaccurate isoform quantification. Some methods also include an Expectation–Maximization algorithm to estimate transcript expression levels when reads map to multiple isoforms or use splice graph reconstruction to infer isoform models based on exon connectivity. Tools that consider isoform-level structure during deduplication are essential to preserve transcriptomic resolution in SCLR-seq data (Table 4).

Several standalone tools and pipelines have been developed specifically for SCLR-seq analysis workflows. Each tool varies in its compatibility with platforms, barcode requirements, and transcript quantification accuracy. While most require a BAM file already tagged with barcodes and UMIs, a few pipelines offer fully integrated workflows from raw FASTQ to expression matrix (Table 2). Some popular pipelines include *FLAMES*, *SiCeLoRe*, *bambu-clump* [98], *nanoseq* [99], and *scNanoGps* [100] (Table 2). *FLAMES* is an end-to-end analysis pipeline for ONT-based SCLR-seq data. It integrates *BLAZE* or *flexiplex* for barcode extraction and uses *bambu* for transcript assembly and *oarfis* [101] for transcript quantification in cells. It generates cell-by-isoform expression matrices and supports multiple UMI-based single-cell protocols. *SiCeLoRe* processes ONT reads by orienting them based on adapter sequences and detecting chimeric cDNAs, which are split and assigned to transcripts. It reconstructs isoforms per cell and produces expression matrices but can be sensitive to read sparsity. *Bambu-clump*, included in the *bambu* package, enables transcript

discovery at the bulk and pseudo-bulk level; it can be used for quantification of fusion transcripts, and it returns full-length read support for all transcripts to support cell-marker validation [98]. Data sequenced using Parse Biosciences' single-cell methods can currently only be preprocessed using Parse's own custom pipeline. For PacBio Iso-Seq data, the pipeline typically involves *Cupcake* to collapse full-length reads and *IsoQuant* to quantify isoform expression. With appropriate barcode tagging, this approach can also be adapted to single-cell analyses.

Standalone tools for isoform discovery and quantification have the advantage of flexibility allowing user customization. *SCOTCH* is a modular toolkit compatible with ONT, PacBio, and a variety of single-cell platforms (e.g. 10x, Parse Biosciences). It reconstructs isoforms directly from long reads and addresses ambiguous mapping challenges through sub-exon identification with dynamic thresholding and read mapping scores [105]. A pseudoalignment for long reads is used in *lr-kallisto*, which is robust to error rates of sequencing platforms [104]. *Isosceles* is an R-based package for ONT data that uses a splice graph-based approach to infer isoform structure, offering high fidelity in transcript resolution [103]. It supports downstream analyses like differential transcript usage (DTU) and clustering at the single-cell, pseudo-bulk, and bulk levels.

Following isoform assembly and quantification, the next step involves classification and quality control of isoforms to remove artifacts and categorize transcript novelty. This is essential to ensure accurate downstream analyses.

Classification of isoforms

Once isoforms have been assembled and quantified, tools such as *SQANTI3* perform detailed classification and filtering. This step uses transcript structural features and external annotations to distinguish high-confidence isoforms from technical artifacts. These artifacts include transcripts with low coverage, RT-switching (reverse transcriptase template switching), and intra-priming (with adenine stretches downstream of the 3' end). *SQANTI3* [113] is currently one of the most comprehensive and widely adopted methods for classifying and filtering isoforms from long-read sequencing data.

Isoform splice junctions are compared with known splice junctions to classify isoforms into the following categories: full splice match, incomplete splice match, novel in catalog (NIC), and novel not in catalog (NNIC). The ordered list of all introns in a read, also called an intron chain, is matched to an annotated transcript. If a read has a novel intron chain but uses known splice sites, it is classified as NIC. If a single splice site on the read differs from known splice sites, it is classified as NNIC [113]. Additionally, if an isoform maps to two different genes, it is classified as a fusion. Depending on which region of the genome a novel transcript maps to, it can be classified as intergenic, genic intron (if located within the boundaries of an annotated intron,) or genic genomic (partial overlap with known gene). A read can also map to the antisense strand of a transcript and is classified as "antisense." As an antisense classified transcript also contains a poly-A tail, it is unlikely that it is a reverse complement. *SQANTI3* can also be used to assess the reliability of isoforms by annotating with publicly available Cap Analysis of Gene Expression peaks and poly(A) sites to filter out PCR artifacts. *SQANTI3* generates an annotated isoform file in gff3 format which is compatible with downstream splicing analysis tools such as *IsoAnnotLite* and *tappAS* [114]. The *IsoAnnot* pipeline integrates coordinate-defined functional annotations at the RNA and protein levels, derived from public databases and sequence-based prediction tools. *tappAS* uses isoform-resolved annotation

Table 2. Bioinformatics tools specific for single-cell long-read sequencing

Tool	Application	Technology	TGS platform	Key features	References
End-to-end pipelines					
FLAMES	Transcript quantification	10× /UMI based	Predominantly for ONT	Isoform discovery, splicing analysis, and mutation detection	[71]
SiCeLoRe	Transcript, gene, and junction quantification	10× /UMI based	Predominantly for ONT	Barcode and UMI identification, isoform quantification, mutation detection	[69]
nanoseq	End to end for SCLR-seq	10×/UMI based	ONT	Quality control, transcript discovery and quantification, differential expression analysis RNA fusion, and RNA modification detection	[99]
EPI2ME Labs / wf-single-cell	ONT's pipeline for SCLR-seq	10×	ONT	Fused read splitting, CB and UMI extraction and correction	github.com/epi2me-labs/wf-single-cell
scycwalker	End to end	10×/UMI based	ONT	Short-read data are optional, pseudo-bulk isoform counts	[102]
Bambu-clump	Transcript discovery	10×/UMI based	ONT	Splice junction-based transcript discovery	[98]
Parse custom pipeline	Demultiplex and preprocessing	Parse single cell	ONT	Deals with multiple barcodes generated due to combinatorial indexing	[88]
Isoseq3	End-to-end pipeline by PacBio	PacBio single cell	PacBio	Tested for single-cell and single-nucleus	[87]
scNanoGPS	Preprocessing quantification and variant calling	10×	ONT		[100]
Identification of cell barcodes and UMI					
scTagger	Match cellular barcodes from long-reads and short-reads	10×	ONT (needs short reads)	Nonzero edit distance matching	[84]
BLAZE	Linux, python	10× /UMI based	Predominantly for ONT	Fast and easy demultiplexing of large-scale SCLR-seq	[85]
Flexplex	Demultiplex and split chimera	10×/UMI based	Predominantly for ONT	A stand-alone tool that identifies cell barcodes and UMIs	[86]
Longcell	UMI deduplication, isoform quantification, and differential splicing analysis	10×/UMI based	Predominantly for ONT	UMI, read mapping correction, and truncation errors without relying on existing isoform annotations	[89]
Standalone isoform quantification					
Isosceles	Isoform assembly and analysis	10× /UMI based	Predominantly for ONT	Isoform detection and quantification across single-cell, pseudo-bulk, and bulk resolution	[103]
Lr-kallisto	Transcript discovery and quantification	10× and Parse	Predominantly for ONT	Pseudoalignment for long-reads	[104]
SCOTCH	Isoform assembly and analysis	10× /UMI based and Parse	Predominantly for ONT	Isoform detection and quantification	[105]
Oarfish	Isoform quantification	UMI based	ONT	Isoform quantification	[101]
Other					
LongSom	Somatic variant calling	10×	Only tested for PacBio	Distinguishes somatic SNVs from noise and germline polymorphisms	[106]
SiSim	Simulate single-cell long-read FASTQ files	10×	ONT	Generates perfect and erroneous reads	github.com/youyupeli/SiSim
AsaruSim	Simulate SCLR-seq data	10×	ONT	Workflow to simulate count matrix and sequences	[107]
CTAT-LR-Fusion	Fusion caller for SCLR-SEQ data	10×	PacBio (tested on simulated ONT)	Detect fusion transcripts	[108]
ScisorWiz	Visualizing differential isoform expression	any	both	Differential isoform expression through various clustering methods	[109]
ScIsoX	Transcriptomic complexity analysis	Any	Both	Metrics that capture transcriptomic complexity across biological scales	[110]
scisorseqr	Differential splicing analysis	10×/UMI based	ONT and PacBio	Barcode deconvolution, mapping, and filtering of high confidence, full-length spliced reads.	[111]
Isopod	DTU analysis	10×/UMI based	ONT and PacBio	Differential transcript usage analysis in R	https://github.com/michael-nakai/isopod

Table 3. Tools dealing with barcodes

Tool	Application	Standalone tool	Isoform clustering	Correct cell barcodes	Correct UMIs	Split chimeras	Key feature	References
BLAZE	Barcode extraction and demultiplexing	Yes	No	Yes	Yes	No	Fast and easy demultiplexing of large scale SCLR-seq	[85]
Flexiplex	Demultiplexing with error-tolerant barcode/ UMI search	Yes	No	Yes	Yes	Yes	Handles split chimeras, error-tolerant demultiplexing	[86]
SiCeLoRe	Barcode/UMI assignment	Part of pipeline		Yes	Yes	Yes	Handles chimeras, accurate assignment of cell barcodes and UMIs	[60]
Longcell	Accurate barcode recovery, iterative clustering of UMIs	Part of pipeline		Yes	Yes	No	Iterative clustering of UMIs, accurate barcode recovery	[89]
scTagger	Match cellular barcodes from long-reads and short-reads	Yes	No	Yes	Yes	No	Nonzero edit distance matching	[84]

All tools are for UMI-based single-cell technologies such as 10x and sequenced on nanopore sequencers.

of coding and non-coding functional domains, motifs, and sites, in combination with novel analysis methods to interrogate different aspects of the functional readout of transcript variants and isoform regulation.

Following isoform classification and filtering, the resulting isoform expression matrices enable many downstream analyses. An immediate next step is usually cell clustering and annotation. Accurate clustering is essential to reveal cellular heterogeneity and assign biological identities to individual cells or cell populations.

Cell clustering and annotation

Read counts must be normalized to be comparable, scaled, and filtered prior to cell clustering. Similar cells are clustered together based on their gene expression profiles or other molecular signatures. Cell clusters are visualized in a 2D map using dimensionality reduction techniques like t-distributed stochastic neighbor embedding [115] or uniform manifold approximation and projection [116]. Cell types are identified by mapping markers expressed in a cell to an annotated cell atlas or a reference set. Assigning biological identity to an individual cell or a distinct cluster can be done manually using domain knowledge or computationally using machine learning models. For example, *Seurat*'s FindMarkers function uses differential expression analysis to identify biomarkers that are defining clusters [117]. It heavily relies on existing biological knowledge of known marker genes and involves subjective decision-making, such as choosing the number of clusters (resolution) [118]. Some examples of automated cell annotation tools include *SingleR* [119] and *scPred* [120].

While tools such as *Seurat*, *Scanpy*, *SingleR*, and *scPred* have been widely adopted for cell clustering and annotation in short-read scRNA-seq datasets, they are not inherently designed to handle isoform-resolved data or take advantage of the full-length transcript information that SCLR-seq provides. These methods typically rely on gene-level expression matrices, which may mask critical isoform-level regulatory differences, particularly in systems where AS plays a major role. When applied to SCLR-seq data, this can lead to suboptimal clustering, as transcripts with distinct functions may be collapsed into a single gene-level expression signal. As such, there is a clear need for new tools or adaptations of existing methods that can fully exploit the isoform resolution offered by SCLR-seq for high-resolution cell-type classification, trajectory inference, and cell-state modeling.

Also, there is currently a lack of standardized reference databases for isoform-level cell-type markers, hindering automated annotation using full-length isoforms. To fully harness the potential of SCLR-seq, there is a need for new computational tools and adaptations of existing ones that operate directly on isoform-resolved expression matrices. Such methods would enable more accurate modeling of cellular heterogeneity, especially in contexts where AS plays a key regulatory role.

Isoform expression and transcriptomic complexity analysis

Once cells have been clustered and annotated, the focus shifts to understanding the transcriptomic diversity that underlies distinct cell types or states. This involves quantifying isoform-level expression and examining patterns of AS, DTU, and other transcriptomic complexities across individual cells. SCLR-seq allows the identification of differential and cell-type-specific splicing events and the quantification of relative isoform abundance. The percent spliced-in (PSI) metric refers to the ratio between reads reflecting the inclusion of a specific sequence element (exon, intron, alternative splice site) and the sum of reads reflecting the inclusion and exclusion of these elements: $\Psi = \text{inclusion reads} / (\text{inclusion reads} + \text{exclusion reads})$. PSI values can elucidate all AS event types, including exon skipping, mutually exclusive exons, alternative 3'/5' splicing, and intron retention. For example, a PSI of 1 indicates constitutive exons that are included in all transcripts and never removed. PSI values below 1 imply reduced inclusion of alternative exons and denote the percentage of transcripts that contain the exon compared to the total transcript population.

Tools to assess differential isoform expression and AS in bulk RNA-seq data such as *SpliceWiz* [120], *rMATS* [121], and *SUPPA2* [122] are well known. Tools specific to assessing differential splicing at the single-cell level include *MARVEL* [123], *BRIE2* [124], *Expedition* [125], and *VALARIE* [126]. Differential isoform expression analysis of SCLR-seq data can be performed with the R-package *scisorseqr* [111], which supports sequencing reads from both Nanopore and PacBio platforms and integrates tools for barcode deconvolution and alignment. However, it is not a pre-processing pipeline and requires that CBs be annotated to cell types before use. There are currently no gold standard methods available for differential splicing and isoform expression analysis specifically for SCLR-seq data.

Table 4. Comparison of tools for isoform discovery and quantification

Tool	Platform	Transcript assembly method	Strengths	Limitations	References
FLAMES	ONT	Semi-supervised isoform discovery and quantification	End-to-end, option to use other tools such as bambu and oarfish	Biased novel isoform, high reads discarded	[71]
SiCeLoRe	ONT	UMI-guided error-correction by consensus generation	Handles chimeras well, end-to-end	High reads discarded, underperforms in highly sparse datasets	[69]
wf-single-cell	ONT	Directed graph algorithm	UMI guided isoform discovery, GUI available for ease of use	Limited benchmarking, memory intensive	github.com/epi2me-labs/wf-single-cell
SCOTCH	ONT/PacBio	Sub-exon identification with dynamic thresholding	Platform-agnostic; no short-read needed	Requires high-quality BAM with tags, newer tool with less validation	[105]
Isosceles	ONT	Splice graph-based isoform reconstruction	High fidelity isoform resolution; DTU support	Specific to ONT; newer tool with less validation, non standard naming of tags	[103]
Lr-kallisto	Predominantly ONT	Pseudo-alignment for long-reads	Not memory intensive, tested for single-nuclei too	Accuracy goes down with high sequencing error rate	[104]
Bambu-clump	ONT	Built on bambu's splice junction matching algorithm	Preprocessing pipeline, internal demultiplexing	Newer tool with less validation	[98]
Cupcake + IsoQuant	PacBio	Collapsing and filtering	Mature workflow for Iso-Seq data	Not single-cell-specific without CB tagging	[87]
LongPreCell (preprocessing to Longcell)	ONT	UMI-based correction of reads by taking consensus	When coupled with IsoQuant gave better outcome in independent study	Newer tool with less validation	[89]
nanoseq	ONT	Integrates previous tools such as bambu, stringtie2	Preprocessing and downstream analysis	Newer tool with less validation as whole pipeline	[99]
oarfish	ONT	Probabilistic method, not UMI aware	Quick, not demanding computationally	Requires preprocessing, no UMI correction	[101]
scNanoGPS	ONT	Deconvolution of error-prone long-reads	Mainly for downstream analysis such as genotype-phenotype correlation	ONT specific, limited benchmark	[100]
bambu/ StringTie2/TALON/FLAIR	ONT/PacBio	Various (annotation-guided, graph-based)	Well-benchmarked for bulk	Cannot natively assign to cells	[92, 93]

For more details authors suggest going through external benchmark papers such as Hamraoui et al. [112].

Differential isoform expression, also referred to as DTU, can reveal cues into development and disease. For instance, Leung *et al.* identified a switch of dominant isoform expression of the *RTN4* gene between adult and fetal brain cortex in humans [127]. *RTN4* encodes a neurite outgrowth inhibitor specific to the central nervous system. DTU analysis reveals the relative abundance of different isoforms of a gene, whereas differential transcript expression (DTE) looks at differences in the combined expression of all isoforms of a gene. DTU informs the proportion of the isoform change between two conditions, whereas DTE informs how much the total expression of the gene has changed between the two conditions. *Sierra* [128] and *DTUrtle* [129] are tools for DTU analysis from scRNA-seq data, while *Isosceles*, *Isopod* (<https://github.com/michael-nakai/isopod>), and *FLAMES* can be used for DTU analysis from single-cell long reads.

SCLR-seq can thoroughly enhance AS analysis, RNA editing, and allelic-specific expression analysis at the single cell level. In addition, long reads cover information on TSS usage and the poly A site usage, patterns of which are evolutionarily conserved and sometimes perturbed in disease [130].

Lastly, transcriptomic complexity analysis describes the systematic examination of global and cell-type-specific isoform expression patterns arising from AS. *ScIsoX* is currently the only framework for multidimensional transcriptomic complexity analysis that uses novel metrics such as (i) inter- and intra-cellular isoform diversity, (ii) intra-cell-type heterogeneity, or (iii) inter-cell-type specificity to capture isoform expression patterns across biological scales. *ScIsoX* operates on isoform count matrices and provides a visualization ecosystem that supports transcriptomic complexity analysis [110].

Identifying fusions and other chimeric RNA

Fusion transcripts arise from gene fusions but can also arise without any DNA rearrangement, e.g. when adjacent genes are transcribed together (transcriptional readthrough) or unconventional splicing events fuse exons from two different transcripts (*trans-splicing*) [131]. Fusion transcripts can also have clinical implications such as the *BCR/ABL1* fusion, a known disease driver in chronic myelogenous leukemia, which is used for diagnosis and is targeted with tyrosine kinase inhibitors [117]. Tools for either scRNA-seq or lRNA-seq data such as JAFFAL [132] and LongShot [133] can be used to identify fusions and other chimeras. SQANTI3 [113] also annotates fusion isoforms. CTAT-LR-fusion is another tool that detects fusion transcripts from lRNA-seq data at both bulk and single-cell levels, with the option of including accompanying short reads in the analysis [108].

Fusions and alternatively spliced transcripts can generate neoepitopes, and their analysis can enhance our understanding of tumor immunogenicity and inform the design of personalized immunotherapies [7, 134, 135]. Kahles et al. identified major histocompatibility complex I (MHC I) epitopes derived from AS events in breast and ovarian cancers using the The Cancer Genome Atlas (TCGA) immunopeptidomics database, finding them to be more abundant than epitopes derived from single-nucleotide variants [136].

Variant calling and allele-specific expression

SCLR-seq provides a unique opportunity to perform somatic variant calling in cells using long reads. Long reads enable the direct observation of multiple variants on the same transcript molecule, facilitating haplotype phasing and detection of SNVs and indels. Allele-specific expression (ASE) in individual cells can reveal how genetic variation impacts specific cell states, lineages, or disease-associated subpopulations, which is very relevant in diseases with clonal mutations like cancer.

Post cell annotation, variants can be called using ClairS [137], sniffles3 [138] or longshot [133], using combined data from clusters representing individual cell types. Variant calling for SCLRS can also be performed via inbuilt options of FLAMES, SiCeLoRe, and the standalone tool LongSom.

By integrating variant information with expression and splicing data, SCLR-seq can be used for analyses such as allele-specific expression mapping, isoform expression quantitative trait locus (eQTL) discovery, and functional interpretation of disease-associated variants in single cells, as described in Zhu et al. [139].

Visualizing differential isoform usage

Visualization is critical for interpreting isoform-level changes across cell types or conditions. Several tools support this step in SCLR-seq workflows. ScisorWiz [109] is a tool specific to SCLR-seq for visualizing differential isoform expression data across multiple clusters. Other general-purpose visualization tools include: *ggsashimi* [127] for visualizing exon-level read coverage and splice junctions, *ggtranscript* [111] an R package for drawing isoform models and comparisons, and Integrative Genomics Viewer, which is widely used for viewing read alignments, including long reads, with support for custom annotations.

Together, these tools facilitate detailed examination of isoform architecture and support discovery of cell-type-specific or disease-associated splicing patterns.

Application of single-cell long-read sequencing in human health research

A deeper understanding of molecular mechanisms at the transcriptome level can be facilitated by identifying splicing patterns at single-cell resolution. SCLR-seq is proving instrumental in studying diverse biological contexts, neurodevelopment, and immune cell diversity, and it is extremely useful in deciphering isoform expression patterns in cancer [66, 67, 76, 134, 140]. Dondi et al. sequenced clinical samples of ovarian cancer and found that mesothelial cells transition into cancer associated fibroblasts, partly via the TGF-beta (transforming growth factor)/miR-29/Collagen axis in omental metastases, by analyzing differential polyadenylation sites and differential isoform usage in collagen-encoding genes *COL1A2*, *COL3A1*, *COL5A2*, and *COL6A1* [66]. They also resolved gene fusion events that were misclassified in matched short-read data, further demonstrating the value of SCLR-seq in oncology and personalized medicine. Li et al. used SCLR-seq to explore the transcriptomic complexity of colorectal cancer, generating a comprehensive isoform atlas [134]. They found novel isoforms in tumor epithelial cells, many arising from complex combinations of splicing events. Some of these isoforms were unique to tumor cells and were experimentally validated. They showed that certain isoforms could generate neoepitopes with strong binding affinities to MHC molecules, suggesting their potential use in developing broadly applicable neoantigen-based cancer vaccines. Yang et al. sequenced single cells from cerebral organoids on both short- and long-read platforms and found that AS events of autistic brains are closer to the progenitor state rather than differentiated neurons, highlighting the importance of cell-type-specific splicing in autism [141].

In a recent study, Dondi et al. developed a method to call *de novo* somatic variants in ovarian cancer, facilitating the study of cancer evolution, clonal heterogeneity, and treatment resistance from SCLR-seq data [106]. Byrne et al. investigated the transcriptional complexity associated with genetic alterations in cancerous cells by associating differentially expressed isoforms to SNVs [68]. Similarly, Wedemeyer's group leveraged SCLR-seq to uncover distinct isoform diversity and transcriptomic profiles in *PIK3CA*-mutant fibroblasts from capillary malformations, providing insights into the molecular consequences of low-frequency mosaic variants in PROS (*PIK3CA*-related overgrowth spectrum) disorders [142].

Beyond somatic variant discovery, SCLR-seq enables ASE analysis and haplotype phasing, allowing the dissection of how cis-regulatory variants affect transcript expression in individual cells. The ability to observe full-length, phased transcripts makes ASE analysis informative for studying allelic regulation, monoallelic expression, and splice site usage in heterogeneous cell populations. In a recent large-scale study, long-read sequencing of B cells from 67 individuals uncovered over 17 000 isoform eQTLs, more than 70% of which were missed by conventional gene-level analysis. Many isoform eQTLs were enriched at splice sites and regulatory histone marks (e.g. H3K36me3, H3K4me1), suggesting that they modulate expression via splicing and transcriptional regulation. Experimental validation using genome editing and minigene splicing assays further confirmed that some functional variants driving isoform expression are overlooked by existing splicing prediction tools [143].

A compelling application of SCLR-seq integrates with spatial transcriptomics to map isoform changes during human brain

development. Traditional spatial transcriptomics, with resolutions around $\sim 60\ \mu\text{m}$, often captures multiple cells per spot. Spl-ISO-seq, a newly developed spatial isoform sequencing method, improves resolution to $\sim 10\ \mu\text{m}$ approaching single-cell granularity to further resolve isoform expression. Spl-ISO-seq, when applied to the human visual cortex across pre- and postpubertal stages, revealed cortical layer and cell-type-specific regulation of AS and poly(A) site usage [144]. A novel application of SCLR-seq involves combining it with chromatin accessibility profiling to better understand how regulatory landscapes influence splicing. In the ScISO-ATAC framework, lrrNA-seq and ATAC-seq are performed simultaneously on nuclei from frozen brain tissue, enabling joint analysis of chromatin accessibility and isoform usage in the same single cells. Hu *et al.* applied this technique to profile human and macaque brain cortices, including tissue from individuals with Alzheimer's disease, to investigate how splicing and chromatin states of cells are regulated. For example, oligodendrocytes exhibit pronounced chromatin and splicing changes in Alzheimer's disease, whereas astrocytes show chromatin dysregulation without major splicing effects [145].

Discussion

Several SCLR-seq datasets have recently become available, reflecting the rapid growth and diversification of this field. Representative studies, including LR-Split-seq [78], ScISO-Seq [67], the PacBio PBMC dataset (<https://downloads.pacbcloud.com/public/dataset/Kinnex-single-cell-RNA/>), and Nanopore-based single-cell transcriptome datasets [68, 146], have profiled complex tissues such as the human and mouse brain, liver, and various cancers. In addition, the recently developed LongBench resource [147] aimed at establishing a standardized reference for cross-platform performance assessment, provides a matched, multi-platform benchmark dataset spanning bulk, single-cell, and single-nucleus transcriptomics. It is from eight human lung cancer cell lines, including ONT (PCR-cDNA and direct RNA) and PacBio Kinnex data alongside Illumina short reads. These datasets employ distinct sequencing platforms, read depths, and library preparation strategies, offering complementary insights into isoform diversity and transcriptome complexity at single-cell resolution. Collectively, they provide a valuable foundation for benchmarking analytical methods and for exploring how lrrNA-seq can uncover cell-type-specific isoforms and aberrant splicing events relevant to human diseases.

SCLR-seq library preparation strategies have a substantial impact on the quality, resolution, and interpretability of data. Key choices in input material, barcoding method, and sequencing platform shape transcript recovery, isoform diversity, and downstream analysis. scRNA-seq captures cytoplasmic transcripts, including mature mRNAs, and is generally richer in isoform content, whereas snRNA-seq captures nuclear RNA, including unspliced or partially spliced transcripts, which can be advantageous for studying nascent transcription or nuclear-retained isoforms, but often underrepresents full-length mRNA.

Barcoding and reverse transcription platform affects both transcript coverage and barcode retention. Droplet-based platforms like 10x Genomics offer high-throughput and standardized barcoding but are limited by 3' or 5' capture bias, often losing full-length information and requiring specialized adaptations to support long-read sequencing. In contrast, plate-based full-length protocols such as SMART-seq3 offer more uniform transcript coverage across the entire mRNA and are better suited for long-read sequencing, albeit at lower throughput and higher cost per

cell. The choice of long-read sequencing platform, PacBio versus ONT, also impacts data quality and read characteristics. PacBio HiFi reads offer higher per-base accuracy ($\geq 99.9\%$), making them ideal for distinguishing similar isoforms or phasing allelic variants, though at lower throughput and higher cost. ONT provides higher throughput, portability, and flexible run configurations, but has higher error rates and greater variability in read length and quality. These trade-offs affect downstream isoform quantification, error correction needs, and feasibility for high-complexity samples.

Together these variables create a matrix of trade-offs between accuracy, throughput, cost, and isoform resolution. Ultimately, the choice of library preparation should be guided by the biological question, target transcript complexity, and available resources.

Challenges and limitations of SCLR-seq

SCLR-seq faces a set of unique computational and technological challenges apart from the obvious limitations of sequencing long reads and sequencing in single cells. When compared to short-read next-generation sequencing (NGS) platforms, TGS platforms generally exhibit lower throughput and higher error rates. Although advances in error correction by ONT and increased throughput by PacBio have greatly improved the efficiency of SCLR-seq, these platforms are still often paired with NGS reads to increase confidence. Higher error rates complicate CB and UMI detection, leading to reduced accuracy in transcript quantification, thus requiring accompanying short-read data.

Single-cell sequencing data are inherently sparse and noisy due to multiple factors. Dropout events occur when gene abundance is too low to be reliably detected, appearing as zeros in the count matrix. The degree of sparsity reflects both true biological variation and methodological noise, which can hinder downstream analyses. In addition to batch effects, technical variability is amplified by amplification bias arising from low starting material. Collectively, these factors can produce incomplete or distorted expression profiles, reducing the reliability of predictions.

Applying long-read sequencing at the single-cell level presents challenges in reconstructing full-length isoforms from noisy, sparse, cell-specific data. Specialized computational tools are needed to address the unique characteristics of SCLR-seq. For example, algorithms must ensure accurate UMI detection to prevent UMI crowding, a phenomenon in which the edit distance between true UMIs decreases as gene expression increases, leading to ambiguous UMI assignment [108]. The use of UMIs helps identify PCR artifacts and reduce amplification bias, making accurate UMI detection essential. Zajac *et al.* compared technical characteristics by sequencing the same single-cell (10x) libraries on Illumina's NovaSeq 6000 and PacBio's Sequel IIe platforms, finding substantial overlap in the results from both [148]. However, genes with high counts in Illumina data were, on average, longer, whereas those with high counts in PacBio data tended to be shorter. In addition to this gene-length bias, Illumina reads mapping to intronic or intergenic regions were often discarded, while PacBio reads provided greater resolution in these regions [148].

Higher throughput can be achieved by sequencing single-cell transcriptomes with short reads rather than long reads at comparable cost. Beyond the expense of single-cell library preparation, research labs face high sequencing costs associated with TGS platforms. Consequently, SCLR-seq experiments remain costly, limiting their accessibility for studies targeting rare cell types or requiring deep sequencing to quantify low-abundance isoforms.

SCLR-seq workflows require significant computational resources for processing and storage. PacBio's Revio sequencer can generate up to 360 Gb of data per day, and ONT's PromethION 48 sequencer can generate a theoretical maximum yield of 14 Tb (when flow-cells at maximum capacity run together for 72 h) [149]. The need for file conversions and preprocessing for downstream analysis further increases the computational data burden, making efficient data storage and computational power critical for successful analysis. Analysis tools are often tailored to a specific experiment type and may not work on a wider range of customized barcodes and sequencing data, specifically, when the structure of the barcode, flanking sequences, and their locations differ from the usual specifications. Much of the available software is complex to set up and install, and its computational requirements in terms of time and memory can limit its practical use when processing the volumes of data now being generated. Moreover, the high costs of commercial kits and capital investment required for SCLR-seq limit the accessibility and flexibility of SCLR-seq. Furthermore, emerging tools have yet to be fully optimized for SCLR-seq analysis, resulting in gaps that limit effective downstream processing.

Conclusions

SCLR-seq overcomes the limitations of conventional short-read single-cell approaches, transforming our ability to study AS at single-cell resolution. Recent advances in both experimental protocols and bioinformatics tools have addressed many initial technical challenges, enabling more accurate and scalable analyses. While short-read sequencing offers higher coverage at lower cost and facilitates barcode deconvolution to improve data reliability, many current workflows still rely on complementary short-read data to boost confidence in cell assignment and transcript quantification. As SCLR-seq technologies continue to mature and become more cost-effective, they may soon be used independently of short reads. Beyond transcriptome reconstruction, SCLR-seq can also integrate well into multi-modal single-cell frameworks where transcript isoform information can be combined with epigenomic, proteomic, or spatial data to provide a more comprehensive view of cellular identity and regulatory mechanisms. This multi-layered perspective will be valuable for dissecting complex diseases, developmental processes, and cell-cell interactions. SCLR-seq holds great promise for uncovering novel molecular insights and deepening our understanding of transcriptomic diversity in health and disease. Continued development of specialized computational pipelines will be critical to fully harness its potential in translational research and precision medicine.

Key Points

- Dysregulation of alternative splicing (AS), a major regulatory mechanism in gene expression, contributes to various human diseases, including cancer.
- This dysregulation has been understood by quantifying transcript isoforms through RNA-seq technologies such as single-cell short-read RNA-seq (scRNA-seq) and bulk lRNA-seq.
- Single-cell long-read RNA sequencing (SCLR-seq) offers a better opportunity to capture isoform diversity than scRNA-seq or lRNA-seq alone.

- This review summarizes recent advances in bioinformatics tools and pipelines for SCLR-seq data analysis and highlights how these developments have been used in human health research.

Author contributions

Saloni Bhatia (Conceptualization, Visualization, Writing—original draft), Matt Field (Supervision, writing—review & editing), Lionel Hebbard (Supervision, Writing—review & editing), and Ulf Schmitz (Conceptualization, Supervision, Writing—review & editing, funding acquisition)

Conflict of interest: None declared.

Funding

This work was supported by James Cook University Postgraduate Research Scholarship (to S.B.); Tropical Australian Academic Health Centre Limited—Research Seed Grant (SF000121 to L.H.). National Health and Medical Research Council Investigator Grants (#5121190 to M.F.; #1196405 to U.S.); Cancer Council NSW Project Grant (RG20-12 to U.S.); Tropical Australian Academic Health Centre (Project Grant SF01124); and the Townsville University Hospital (Grants #THHSERTA RPG05 2024, #THHSERTA RPG15 2024, #THHSERTA RCG05 2024, and #RPG09_2025).

Data availability

There are no new data associated with this article.

References

1. Vaquero-Garcia J, Barrera A, Gazzara MR. et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *eLife* 2016;5:e11752. <https://doi.org/10.7554/eLife.11752>
2. Weyn-Vanhenhenryck SM, Feng H, Ustianenko D. et al. Precise temporal regulation of alternative splicing during neural development. *Nat Commun* 2018;9:2189–17. <https://doi.org/10.1038/s41467-018-04559-0>
3. DeBoever C, Ghia EM, Shepard PJ. et al. Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. *PLoS Comput Biol* 2015; 11:e1004105–5. <https://doi.org/10.1371/journal.pcbi.1004105>
4. Jiang W, Chen L. Alternative splicing: human disease and quantitative analysis from high-throughput sequencing. *Comput Struct Biotechnol J* 2021;19:183–95. <https://doi.org/10.1016/j.csbj.2020.12.009>
5. Xu J, Gong B, Wu L. et al. Comprehensive assessments of RNA-seq by the SEQC consortium: FDA-led efforts advance precision medicine. *Pharmaceutics* 2016;8:8. <https://doi.org/10.3390/pharmaceutics8010008>
6. Zhao J, Guo S, Schrodri SJ. et al. Molecular and cellular heterogeneity in rheumatoid arthritis: mechanisms and clinical implications. *Front Immunol* 2021;12:790122. <https://doi.org/10.3389/fimmu.2021.790122>
7. Dorney R, Dhungel BP, Rasko JEJ. et al. Recent advances in cancer fusion transcript detection. *Brief Bioinform* 2023;24:bbac519.

8. Marinov GK, Williams BA, McCue K. et al. From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res* 2014;**24**:496–510. <https://doi.org/10.1101/gr.161034.113>
9. Shalek AK, Satija R, Adiconis X. et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature (London)* 2013;**498**:236–40. <https://doi.org/10.1038/nature12172>
10. Yap K, Makeyev EV. Functional impact of splice isoform diversity in individual cells. *Biochem Soc Trans* 2016;**44**:1079–85. <https://doi.org/10.1042/BST20160103>
11. Feng H, Moakley DF, Chen S. et al. Complexity and graded regulation of neuronal cell-type-specific alternative splicing revealed by single-cell RNA sequencing. *Proc Natl Acad Sci* 2021;**118**:1–12. <https://doi.org/10.1073/pnas.2013056118>
12. Jiang H, Yu D, Yang P. et al. Revealing the transcriptional heterogeneity of organ-specific metastasis in human gastric cancer using single-cell RNA sequencing. *Clin Transl Med* 2022;**12**:e730. <https://doi.org/10.1002/ctm2.730>
13. Lai H, Cheng X, Liu Q. et al. Single-cell RNA sequencing reveals the epithelial cell heterogeneity and invasive subpopulation in human bladder cancer. *Int J Cancer* 2021;**149**:2099–115. <https://doi.org/10.1002/ijc.33794>
14. Maynard A, McCoach CE, Rotow JK. et al. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* 2020;**182**:1232–1251.e22. <https://doi.org/10.1016/j.cell.2020.07.017>
15. Glinos DA, Garborcauskas G, Hoffman P. et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature (London)* 2022;**608**:353–9. <https://doi.org/10.1038/s41586-022-05035-y>
16. Alfonso-Gonzalez C, Legnini I, Holec S. et al. Sites of transcription initiation drive mRNA isoform selection. *Cell* 2023;**186**:2438–2455.e22. <https://doi.org/10.1016/j.cell.2023.04.012>
17. Anvar SY, Allard G, Tseng E. et al. Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biol* 2018;**19**:46–6. <https://doi.org/10.1186/s13059-018-1418-0>
18. Wu S, Schmitz U. Single-cell and long-read sequencing to enhance modelling of splicing and cell-fate determination. *Comput Struct Biotechnol J* 2023;**21**:2373–80. <https://doi.org/10.1016/j.csbj.2023.03.023>
19. Belchikov N, Hsu J, Li XJ. et al. Understanding isoform expression by pairing long-read sequencing with single-cell and spatial transcriptomics. *Genome Res* 2024;**34**:1735–46. <https://doi.org/10.1101/gr.279640.124>
20. Wang ET, Sandberg R, Luo S. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008;**456**:470–6. <https://doi.org/10.1038/nature07509>
21. Raj B, Blencowe BJ. Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles. *Neuron* 2015;**87**:14–27.
22. Pan Q, Shai O, Lee LJ. et al. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 2008;**40**:1413–5. <https://doi.org/10.1038/ng.259>
23. Schertzer MD, Park SH, Su J. et al. Perplexity as a metric for isoform diversity in the human transcriptome. *bioRxiv*. 2025; preprint: 2025.07.02.662769. <https://doi.org/10.1101/2025.07.02.662769>
24. Schwerk C, Schulze-Osthoff K. Regulation of apoptosis by alternative pre-mRNA splicing. *Mol Cell* 2005;**19**:1–13. <https://doi.org/10.1016/j.molcel.2005.05.026>
25. Stevens M, Oltean S. Modulation of the apoptosis gene Bcl-x function through alternative splicing. *Front Genet* 2019;**10**:804–4. <https://doi.org/10.3389/fgene.2019.00804>
26. Wong JLL, Ritchie W, Ebner OA. et al. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 2013;**154**:583–95. <https://doi.org/10.1016/j.cell.2013.06.052>
27. Wong JLL, Schmitz U. Intron retention: importance, challenges, and opportunities. *Trends Genet* 2022;**38**:789–92. <https://doi.org/10.1016/j.tig.2022.03.017>
28. Lim CS, Wardell SJ, Kleffmann T. et al. The exon-intron gene structure upstream of the initiation codon predicts translation efficiency. *Nucleic Acids Res* 2018;**46**:4575–91. <https://doi.org/10.1093/nar/gky282>
29. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet* 2016;**17**:19–32. <https://doi.org/10.1038/nrg.2015.3>
30. Wang R, Helbig I, Edmondson AC. et al. Splicing defects in rare diseases: transcriptomics and machine learning strategies towards genetic diagnosis. *Brief Bioinform* 2023;**24**:bbad284.
31. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet* 2018;**59**:253–68. <https://doi.org/10.1007/s13353-018-0444-7>
32. Vanichkina DP, Schmitz U, Wong JLL. et al. Challenges in defining the role of intron retention in normal biology and disease. *Semin Cell Dev Biol* 2018;**75**:40–9. <https://doi.org/10.1016/j.semcdb.2017.07.030>
33. Li D, Wang Q, Bayat A. et al. Spliceosome malfunction causes neurodevelopmental disorders with overlapping features. *J Clin Invest* 2024;**134**:1–16. <https://doi.org/10.1172/JCI171235>
34. Gandal MJ, Zhang P, Hadjimichael E. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* 2018;**362**:eaat8127.
35. Bradley RK, Anczuków O. RNA splicing dysregulation and the hallmarks of cancer. *Nat Rev Cancer* 2023;**23**:135–55. <https://doi.org/10.1038/s41568-022-00541-7>
36. Petrova V, Song R, Nordström KJV. et al. Increased chromatin accessibility facilitates intron retention in specific cell differentiation states. *Nucleic Acids Res* 2022;**50**:11563–79. <https://doi.org/10.1093/nar/gkac994>
37. Sciarillo R, Wojtuszkiewicz A, Assaraf YG. et al. The role of alternative splicing in cancer: from oncogenesis to drug resistance. *Drug Resist Updat* 2020;**53**:100728–8. <https://doi.org/10.1016/j.drug.2020.100728>
38. Verdile V, Palombo R, Ferrante G. et al. Dysregulation of alternative splicing underlies synaptic defects in familial amyotrophic lateral sclerosis. *Prog Neurobiol* 2023;**231**:102529–9. <https://doi.org/10.1016/j.pneurobio.2023.102529>
39. Puri S, Maachi H, Nair G. et al. Sox9 regulates alternative splicing and pancreatic beta cell function. *Nat Commun* 2024;**15**:588–8. <https://doi.org/10.1038/s41467-023-44384-8>
40. Nitschke L, Hu R-C, Miller AN. et al. Rescue of Scn5a missplicing does not improve the structural and functional heart defects of a DM1 heart mouse model. *Hum Mol Genet* 2024;**33**:1789–99. <https://doi.org/10.1093/hmg/ddae117>
41. Robert F, Pelletier J. Exploring the impact of single-nucleotide polymorphisms on translation. *Front Genet* 2018;**9**:507. <https://doi.org/10.3389/fgene.2018.00507>
42. Schulz J, Mah N, Neuenschwander M. et al. Loss-of-function uORF mutations in human malignancies. *Sci Rep* 2018;**8**:2395–10. <https://doi.org/10.1038/s41598-018-19201-8>
43. McAlinden A, Majava M, Bishop PN. et al. Missense and non-sense mutations in the alternatively-spliced exon 2 of COL2A1 cause the ocular variant of stickler syndrome. *Hum Mutat* 2008;**29**:83–90. <https://doi.org/10.1002/humu.20603>

44. Bhadra M, Howell P, Dutta S. et al. Alternative splicing in aging and longevity. *Hum Genet* 2020;**139**:357–69. <https://doi.org/10.1007/s00439-019-02094-6>
45. Nikom D, Zheng S. Alternative splicing in neurodegenerative disease and the promise of RNA therapies. *Nat Rev Neurosci* 2023;**24**:457–73. <https://doi.org/10.1038/s41583-023-00717-6>
46. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. *Oncogene* 2014;**33**:5311–8. <https://doi.org/10.1038/ncr.2013.533>
47. Tian N, Li J, Shi J. et al. From general aberrant alternative splicing in cancers and its therapeutic application to the discovery of an oncogenic DMTF1 isoform. *Int J Mol Sci* 2017;**18**:191. <https://doi.org/10.3390/ijms18030191>
48. Chen C, Zhao S, Karnad A. et al. The biology and role of CD44 in cancer progression: therapeutic implications. *J Hematol Oncol* 2018;**11**:64. <https://doi.org/10.1186/s13045-018-0605-5>
49. Escobar-Hoyos LF, Penson A, Kannan R. et al. Altered RNA splicing by mutant p53 activates oncogenic RAS signaling in pancreatic cancer. *Cancer Cell* 2020;**38**:198–211.e8. <https://doi.org/10.1016/j.ccell.2020.05.010>
50. Cheng R, Xiao L, Zhou W. et al. A pan-cancer analysis of alternative splicing of splicing factors in 6904 patients. *Oncogene* 2021;**40**:5441–50. <https://doi.org/10.1038/s41388-021-01947-7>
51. Sen S, Langiewicz M, Jumaa H. et al. Deletion of serine/arginine-rich splicing factor 3 in hepatocytes predisposes to hepatocellular carcinoma in mice. *Hepatology* 2015;**61**:171–83.
52. Su M, Fleischer T, Grosheva I. et al. Targeting SRSF2 mutations in leukemia with RKI-1447: a strategy to impair cellular division and nuclear structure. *iScience* 2024;**27**:109443. <https://doi.org/10.1016/j.isci.2024.109443>
53. Alfonso-Gonzalez C, Hilgers V. Elucidating the coordination of RNA processing using short-read and long-read RNA-sequencing methods. *Nat Rev Mol Cell Biol* 2025;1–19. <https://doi.org/10.1038/s41580-025-00895-4>
54. Hong D, Jeong S. 3'UTR diversity: expanding repertoire of RNA alterations in human mRNAs. *Mol Cells* 2023;**46**:48–56. <https://doi.org/10.14348/molcells.2023.0003>
55. He PC, He C. m6A RNA methylation: from mechanisms to therapeutic potential. *EMBO J* 2021;**40**:e105977. <https://doi.org/10.15252/embj.2020105977>
56. Walkley CR, Li JB. Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. *Genome Biol* 2017;**18**:205–5. <https://doi.org/10.1186/s13059-017-1347-3>
57. Mukherjee S, Heng HH, Frenkel-Morgenstern M. Emerging role of chimeric RNAs in cell plasticity and adaptive evolution of cancer cells. *Cancers* 2021;**13**:4328. <https://doi.org/10.3390/cancers13174328>
58. Li Y, Liu Y, Xie Y. et al. Long-read RNA sequencing enables full-length chimeric transcript annotation of transposable elements in lung adenocarcinoma. *BMC Cancer* 2025;**25**:482–16. <https://doi.org/10.1186/s12885-025-13888-5>
59. Rojo-Carrillo JJ, Garrido-Rodríguez P, Llamas-López M. et al. Landscape of antisense genes in the human genome and identification of new human hepatic antisense RNAs by long-read sequencing. *BMC Genomics* 2024;**25**:1148–20. <https://doi.org/10.1186/s12864-024-11017-3>
60. Oehler JB, Wright H, Stark Z. et al. The application of long-read sequencing in clinical settings. *Hum Genomics* 2023;**17**:73–13. <https://doi.org/10.1186/s40246-023-00522-3>
61. Roberts RJ, Carneiro MO, Schatz MC. Advantages of SMRT sequencing. *Genome Biol* 2013;**14**:405–5. <https://doi.org/10.1186/gb-2013-14-6-405>
62. Espinosa E, Bautista R, Larrosa R. et al. Advancements in long-read genome sequencing technologies and algorithms. *Genomics* 2024;**116**:110842.
63. Al'Khafaji AM, Smith JT, Garimella KV. et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat Biotechnol* 2024;**42**:582–6. <https://doi.org/10.1038/s41587-023-01815-7>
64. Philpott M, Watson J, Thakurta A. et al. Nanopore sequencing of single-cell transcriptomes with scCOLOR-seq. *Nat Biotechnol* 2021;**39**:1517–20. <https://doi.org/10.1038/s41587-021-00965-w>
65. Scarano C, Veneruso I, De Simone RR. et al. The third-generation sequencing challenge: novel insights for the Omic sciences. *Biomolecules* 2024;**14**:568. <https://doi.org/10.3390/biom14050568>
66. Dondi A, Lischetti U, Jacob F. et al. Detection of isoforms and genomic alterations by high-throughput full-length single-cell RNA sequencing in ovarian cancer. *Nat Commun* 2023;**14**:7780–19. <https://doi.org/10.1038/s41467-023-43387-9>
67. Gupta I, Collier PG, Haase B. et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat Biotechnol* 2018;**36**:1197–202. <https://doi.org/10.1038/nbt.4259>
68. Byrne A, Le D, Sereti K. et al. Single-cell long-read targeted sequencing reveals transcriptional variation in ovarian cancer. *Nat Commun* 2024;**15**:6916–3. <https://doi.org/10.1038/s41467-024-51252-6>
69. Lebrigand K, Magnone V, Barbry P. et al. High throughput error corrected nanopore single cell transcriptome sequencing. *Nat Commun* 2020;**11**:1–8. <https://doi.org/10.1038/s41467-020-17800-6>
70. Wang X, He Y, Zhang Q. et al. Direct comparative analyses of 10X genomics chromium and Smart-seq2. *Genomics Proteomics Bioinformatics* 2021;**19**:253–66. <https://doi.org/10.1016/j.gpb.2020.02.005>
71. Tian L, Jabbari JS, Thijssen R. et al. Comprehensive characterization of single-cell full-length isoforms in human and mouse with long-read sequencing. *Genome Biol* 2021;**22**:310–0. <https://doi.org/10.1186/s13059-021-02525-6>
72. Singh M, Al-Eryani G, Carswell S. et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat Commun* 2019;**10**:3120–13. <https://doi.org/10.1038/s41467-019-11049-4>
73. Peng H, Jabbari JS, Tian L. et al. Single-cell rapid capture hybridization sequencing reliably detects isoform usage and coding mutations in targeted genes. *Genome Res* 2025;**35**:942–55. <https://doi.org/10.1101/gr.279322.124>
74. Grabek J, Straube J, Cooper L. et al. Single cell long-read genotyping of transcripts reveals discrete mechanisms of clonal evolution in post-myeloproliferative neoplasm acute myeloid leukemia. *bioRxiv*. 2025; preprint: 2025.08.18.670417. <https://doi.org/10.1101/2025.08.18.670417>
75. Picelli S, Faridani OR, Björklund ÅK. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 2014;**9**:171–81. <https://doi.org/10.1038/nprot.2014.006>
76. Volden R, Palmer T, Byrne A. et al. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc Natl Acad Sci U S A* 2018;**115**:9726–31. <https://doi.org/10.1073/pnas.1806447115>
77. Liao Y, Liu Z, Zhang Y. et al. High-throughput and high-sensitivity full-length single-cell RNA-seq analysis on third-generation sequencing platform. *Cell Discov* 2023;**9**:5–5. <https://doi.org/10.1038/s41421-022-00500-4>
78. Rebboah E, Reese F, Williams K. et al. Mapping and modeling the genomic basis of differential RNA isoform expression at single-cell resolution with LR-split-seq. *Genome Biol* 2021;**22**:286–6. <https://doi.org/10.1186/s13059-021-02505-w>

79. Parse Biosciences. Multi-omics approach for near-full-length human iPSC transcriptomes in cardiomyocyte models. Seattle (WA): Parse Biosciences; [cited 2025 Dec 4]. Available from <https://www.parsebiosciences.com/customer-datasets/multi-omics-approach-for-near-full-length-human-ipsc-transcriptomes-in-cardiomyocyte-models/>
80. Gupta P, O'Neill H, Wolvetang EJ. et al. Advances in single-cell long-read sequencing technologies. *NAR Genom Bioinform* 2024;**6**:lqae047. <https://doi.org/10.1093/nargab/lqae047>
81. Zappia L, Phipson B, Oshlack A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput Biol* 2018;**14**:e1006245–5. <https://doi.org/10.1371/journal.pcbi.1006245>
82. Amarasinghe SL, Ritchie ME, Gouil Q. long-read-tools.org: an interactive catalogue of analysis methods for long-read sequencing data. *Gigascience* 2021;**10**:giab003. <https://doi.org/10.1093/gigascience/giab003>
83. Gamaarachchi H, Samarakoon H, Jenner SP. et al. Fast nanopore sequencing data analysis with SLOW5. *Nat Biotechnol* 2022;**40**:1026–9. <https://doi.org/10.1038/s41587-021-01147-4>
84. Ebrahimi G, Orabi B, Robinson M. et al. Fast and accurate matching of cellular barcodes across short-reads and long-reads of single-cell RNA-seq experiments. *iScience* 2022;**25**:104530. <https://doi.org/10.1016/j.isci.2022.104530>
85. You Y, Prawer YDJ, De Paoli-Iseppi R. et al. Identification of cell barcodes from long-read single-cell RNA-seq with BLAZE. *Genome Biol* 2023;**24**:66. <https://doi.org/10.1186/s13059-023-02907-y>
86. Cheng O, Ling MH, Wang C. et al. Flexiplex: a versatile demultiplexer and search tool for omics data. *Bioinformatics* 2024;**40**:btac102. <https://doi.org/10.1093/bioinformatics/btac102>
87. Guizard S, Miedzinska K, Smith J. et al. Nf-core/isoseq: simple gene and isoform annotation with PacBio iso-Seq long-read sequencing. *Bioinformatics* 2023;**39**:btad150. <https://doi.org/10.1093/bioinformatics/btad150>
88. Parse Biosciences. Technology overview. Seattle, USA: Parse Biosciences [cited 2025 Dec 4]. Available from <https://www.parsebiosciences.com/technology/>
89. Fu Y, Kim H, Roy S. et al. Single cell and spatial alternative splicing analysis with nanopore long read sequencing. *Nat Commun* 2025;**16**:6654. <https://doi.org/10.1038/s41467-025-60902-2>
90. Deng E, Shen Q, Zhang J. et al. Systematic evaluation of single-cell RNA-seq analyses performance based on long-read sequencing platforms. *J Adv Res* 2024;**71**:141–53. <https://doi.org/10.2991/978-94-6463-564-5>
91. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100. <https://doi.org/10.1093/bioinformatics/bty191>
92. Chen Y, Sim A, Wan YK. et al. Context-aware transcript quantification from long-read RNA-seq data with Bambu. *Nat Methods* 2023;**20**:1187–95. <https://doi.org/10.1038/s41592-023-01908-w>
93. Kovaka S, Zimin AV, Pertea GM. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* 2019;**20**:278. <https://doi.org/10.1186/s13059-019-1910-1>
94. Prjibelski AD, Mikheenko A, Joglekar A. et al. Accurate isoform discovery with IsoQuant using long reads. *Nat Biotechnol* 2023;**41**:915–8. <https://doi.org/10.1038/s41587-022-01565-y>
95. Su Y, Yu Z, Jin S. et al. Comprehensive assessment of mRNA isoform detection methods for long-read sequencing data. *Nat Commun* 2024;**15**:3972. <https://doi.org/10.1038/s41467-024-48117-3>
96. Smith T, Heger A, Sudbery I. UMI-tools: Modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res* 2017;**27**:491–9. <https://doi.org/10.1101/gr.209601.116>
97. Srivastava A, Malik L, Smith T. et al. Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. *Genome Biol* 2019;**20**:65. <https://doi.org/10.1186/s13059-019-1670-y>
98. Sim A, Ling MH, Chen Y. et al. Isoform-level discovery, quantification and fusion analysis from single-cell and spatial long-read RNA-seq data with Bambu-clump. *bioRxiv* 2024.12.30.630828 [preprint]. <https://doi.org/10.1101/2024.12.30.630828>
99. Chen Y, Davidson NM, Wan YK. et al. A systematic benchmark of nanopore long-read RNA sequencing for transcript-level analysis in human cell lines. *Nat Methods* 2025;**22**:801–12. <https://doi.org/10.1038/s41592-025-02623-4>
100. Shiao C-K, Lu L, Kieser R. et al. High throughput single cell long-read sequencing analyses of same-cell genotypes and phenotypes in human tumors. *Nat Commun* 2023;**14**:4124–4. <https://doi.org/10.1038/s41467-023-39813-7>
101. Zare Jousheghani Z, Singh NP, Patro R. Oarfish: enhanced probabilistic modeling leads to improved accuracy in long read transcriptome quantification. *Bioinformatics* 2025;**41**:i304–13.
102. De Rijk P, Watzeels T, Küçükali F. Scywalker: scalable end-to-end data analysis workflow for long-read single-cell transcriptome sequencing. *Bioinformatics* 2024;**40**.
103. Kabza M, Ritter A, Byrne A. et al. Accurate long-read transcript discovery and quantification at single-cell, pseudo-bulk and bulk resolution with isosecles. *Nat Commun* 2024;**15**:7316–2.
104. Loving RK, Sullivan DK, Boeshagi AS. et al. Long-read sequencing transcriptome quantification with lr-kallisto. *bioRxiv* 2024.07.19.604364 [preprint]. <https://doi.org/10.1101/2024.07.19.604364>
105. Xu Z, Hui-Qi Q, Kao C. et al. Single-cell omics for transcriptome CHaracterization (SCOTCH): isoform-level characterization of gene expression through long-read single-cell RNA sequencing. *bioRxiv* 2024.04.29.590597 [preprint]. <https://doi.org/10.1101/2024.04.29.590597>
106. Dondi A, Borgsmüller N, Ferreira PF. et al. De novo detection of somatic variants in high-quality long-read single-cell RNA sequencing data. *Genome Res* 2025;**35**:900–13. <https://doi.org/10.1101/gr.279281.124>
107. Hamraoui A, Jourden L, Thomas-Chollier M. AsaruSim: a single-cell and spatial RNA-Seq Nanopore long-reads simulation workflow. *Bioinformatics* 2025;**41**.
108. Qin Q, Popic V, Wienand K. et al. Accurate fusion transcript identification from long- and short-read isoform sequencing at bulk or single-cell resolution. *Genome Res* 2025;**35**:967–86.
109. Stein AN, Joglekar A, Poon C-L. et al. ScisorWiz: Visualizing differential isoform expression in single-cell long-read data. *Bioinformatics* 2022;**38**:3474–6.
110. Wu S, Schmitz U. ScIsoX: a multidimensional framework for measuring isoform-level transcriptomic complexity in single cells. *Genome Biol* 2025;**26**:289.
111. Joglekar A, Prjibelski A, Mahfouz A. et al. A spatially resolved brain region- and cell type-specific isoform atlas of the post-natal mouse brain. *Nat Commun* 2021;**12**:463–3.
112. Hamraoui A, Onfroy A, Senamaud-Beaufort C. et al. A systematic benchmark of bioinformatics methods for single-cell and spatial RNA-seq nanopore long-read data. *bioRxiv* 2025.07.21.665920 [preprint]. <https://doi.org/10.1101/2025.07.21.665920>
113. Pardo-Palacios FJ, Arzalluz-Luque A, Kondratova L. et al. SQANTI3: curation of long-read transcriptomes for

- accurate identification of known and novel isoforms. *Nat Methods* 2024;**21**:793–7.
114. de la Fuente L, Arzalluz-Luque Á, Tardáguila M. et al. tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing. *Genome Biol* 2020;**21**:119–9.
 115. van der Maaten L, Hinton G. Visualizing non-metric similarities in multiple maps. *Mach Learn* 2012;**87**:33–55.
 116. Becht E, McInnes L, Healy J. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2019;**37**:38–44.
 117. Hao Y, Stuart T, Kowalski MH. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 2024;**42**:293–304.
 118. Cheng C, Chen W, Jin H. et al. A review of single-cell RNA-Seq annotation, integration, and cell-cell communication. *Cells* 2023;**12**:1970.
 119. Aran D, Looney AP, Liu L. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 2019;**20**:163–72.
 120. Wong ACH, Wong JLL, Rasko JEJ. et al. SpliceWiz: interactive analysis and visualization of alternative splicing in R. *Brief Bioinform* 2023;**25**:bbad468.
 121. Shen S, Park JW, Lu Z-X. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* 2014;**111**:E5593–601.
 122. Trincado JL, Entizne JC, Hysenaj G. et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* 2018;**19**:40.
 123. Wen WX, Mead AJ, Thongjuea S. MARVEL: an integrated alternative splicing analysis platform for single-cell RNA sequencing data. *Nucleic Acids Res* 2023;**51**:e29–9.
 124. Huang Y, Sanguinetti G. BRIE2: computational identification of splicing phenotypes from single-cell transcriptomic experiments. *Genome Biol* 2021;**22**:251–1.
 125. Song Y, Botvinnik OB, Lovci MT. et al. Single-cell alternative splicing analysis with expedition reveals splicing dynamics during neuron differentiation. *Mol Cell* 2017;**67**:148–161.e145.
 126. Wen WX, Mead AJ, Thongjuea S. VALERIE: visual-based inspection of alternative splicing events at single-cell resolution. *PLoS Comput Biol* 2020;**16**:e1008195–5.
 127. Leung SK, Jeffries AR, Castanho I. et al. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Rep* 2021;**37**:110022.
 128. Patrick R, Humphreys DT, Janbandhu V. et al. Sierra: discovery of differential transcript usage from polyA-captured single-cell RNA-seq data. *Genome Biol* 2020;**21**:167.
 129. Tekath T, Dugas M. Differential transcript usage analysis of bulk and single-cell RNA-seq data with DTUrtle. *Bioinformatics* 2021;**37**:3781–7.
 130. Xiong J, Jiang X, Ditsiou A. et al. Predominant patterns of splicing evolution on human, chimpanzee and macaque evolutionary lineages. *Hum Mol Genet* 2018;**27**:1474–85.
 131. Li H, Wang J, Mor G. et al. Neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* 2008;**321**:1357–61.
 132. Davidson NM, Chen Y, Sadras T. et al. JAFFAL: detecting fusion genes with long-read transcriptome sequencing. *Genome Biol* 2022;**23**:10.
 133. Edge P, Bansal V. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat Commun* 2019;**10**:4660.
 134. Li Z, Zhang B, Chan JJ. et al. An isoform-resolution transcriptomic atlas of colorectal cancer from long-read single-cell sequencing. *Cell Genom* 2024;**4**:100641.
 135. Smart AC, Margolis CA, Pimentel H. et al. Intron retention is a source of neopeptides in cancer. *Nat Biotechnol* 2018;**36**:1056–8.
 136. Kahles A, Lehmann KV, Toussaint NC. et al. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* 2018;**34**:211–224.e216.
 137. Chen L, Zheng Z, Su J. et al. ClairS-TO: a deep-learning method for long-read tumor-only somatic small variant calling. *Nat Commun* 2025;**16**:9630. <https://doi.org/10.1038/s41467-025-64547-z>
 138. Smolka M, Paulin LF, Grochowski CM. et al. Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol* 2024;**42**:1571–80.
 139. Zhu F, Liu Y, Lei J. et al. Single-cell eQTL mapping of human endogenous retroviruses reveals cell type-specific genetic regulation in autoimmune diseases. *Nat Commun* 2025;**16**:7534.
 140. Wu X, Lu M, Yun D. et al. Long-read single-cell sequencing reveals the transcriptional landscape of spermatogenesis in obstructive azoospermia and Sertoli cell-only patients. *QJM* 2024;**117**:422–35.
 141. Yang Y, Yang R, Kang B. et al. Single-cell long-read sequencing in human cerebral organoids uncovers cell-type-specific and autism-associated exons. *Cell Rep* 2023;**42**:113335.
 142. Wedemeyer MA, Ding T, Garfinkle EAR. et al. Defining the transcriptome of PIK3CA-altered cells in a human capillary malformation using single cell long-read sequencing. *Sci Rep* 2024;**14**:25440.
 143. Nagura Y, Shimada M, Kuribayashi R. et al. Long-read sequencing reveals novel isoform-specific eQTLs and regulatory mechanisms of isoform expression in human B cells. *Genome Biol* 2025;**26**:110.
 144. Foord C, Pribelski AD, Hu W. et al. A spatial long-read approach at near-single-cell resolution reveals developmental regulation of splicing and polyadenylation sites in distinct cortical layers and cell types. *Nat Commun* 2025;**16**:8093.
 145. Hu W, Foord C, Hsu J. et al. Combined single-cell profiling of chromatin-transcriptome and splicing across brain cell types, regions and disease state. *Nat Biotechnol* 2025;**22**:10.1038/s41587-025-02734-5. <https://doi.org/10.1038/s41587-025-02734-5>
 146. Joglekar A, Hu W, Zhang B. et al. Single-cell long-read sequencing-based mapping reveals specialized splicing patterns in developing and adult mouse and human brain. *Nat Neurosci* 2024;**27**:1051–63.
 147. You Y, Solano A, Lancaster J. et al. Benchmarking long-read RNA-sequencing technologies with LongBench: a cross-platform reference dataset profiling cancer cell lines with bulk and single-cell approaches. *bioRxiv* 2025.09.11.675724 [preprint]. <https://doi.org/10.1101/2025.09.11.675724>
 148. Zajac N, Zhang Q, Bratus-Neuenschwander A. et al. Comparison of single-cell long-read and short-read transcriptome sequencing via cDNA molecule matching: quality evaluation of the MAS-ISO-seq approach. *NAR Genom Bioinform* 2025;**7**:lqaf089.
 149. van Dijk EL, Jaszczyszyn Y, Naquin D. et al. The third revolution in sequencing technology. *Trends Genet* 2018;**34**:666–81.