Briefings in Bioinformatics, 2026, 27, bbag323
https://doi.org/10.1093/bib/bbag323
Published: 17 June 2026
Review

OXFORD

Current trends and challenges in deciphering single molecule resolution maps of single cell transcriptomes

David Schaeper[1], Upol Chowdhury[1], Sarath Chandra Janga[1,2,3,*]

[1]Department of Biomedical Engineering and Informatics, Luddy School of Informatics, Computing and Engineering, Indiana University Indianapolis (IU Indianapolis), 535 West Michigan Street, Indianapolis, IN 46202, United States
[2]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Medical Research and Library Building, 975 West Walnut Street, Indianapolis, IN 46202, United States
[3]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Health Information and Translational Sciences (HITS) Building, Suite 5000, 410 West 10th Street, Indianapolis, IN 46202, United States
*Corresponding author. Department of Biomedical Engineering and Informatics, Luddy School of Informatics, Computing and Engineering, IU Indianapolis, Informatics and Communications Technology Complex, IT479, 535 West Michigan Street, Indianapolis, IN 46202, United States. E-mail: scjanga@iu.edu

Abstract

Advancements in single-cell RNA sequencing (scRNA-seq) techniques have expanded the study of cellular heterogeneity and transcriptional dynamics. Early methods relied on manual cell isolation followed by barcode introduction, but subsequent approaches integrated automated cell isolation with cellular barcoding to increase throughput. While most current single-cell RNA-seq methods aim to capture transcripts at the single-cell level in a high-throughput manner using short-read sequencing, such efforts frequently prevent assignment of full-length transcripts to individual cells, limiting insight into isoform diversity and complete mutational profiles. Recent advances in long-read sequencing accuracy are starting to enable integration of full-length transcript coverage with high-throughput barcoding. This review traces the evolution of scRNA-seq from early manual isolation methods to today's high-throughput short-read droplet- and combinatorial barcoding-based platforms. Then, the review discusses recent advances stemming from the adaptation of high-throughput scRNA-seq protocols for use with long-read sequencing and addresses key challenges such as accurate barcode identification despite lower base-calling accuracy and efforts to compensate for reduced throughput relative to short-read technologies. In parallel, the review highlights the development of computational tools tailored to long-read scRNA-seq, including methods for cell barcode and unique molecular index recovery, variant detection, and complete end-to-end workflows, emphasizing both their shared and unique advantages. Finally, applications of long-read scRNA-seq are shown to provide novel insights, spanning cancer genomics, neurology, early development, and disease contexts. By integrating technical, computational, and biological perspectives, the transformative potential of long-read scRNA-seq is shown, advancing our understanding of cellular heterogeneity.

Keywords **single molecule, long-read sequencing, single cell transcriptomes, computational workflows, precision medicine, isoform resolution**

Introduction

Traditionally, profiling the RNA of an organism was performed using bulk RNA sequencing. However, since bulk RNA sequencing and similar protocols aggregate many cells into a single experiment, while providing valuable transcriptomic information of the entire sample, key insights into the RNA expression of different cell types and states are not captured. Single-cell RNA sequencing (scRNA-seq) techniques were developed to address this limitation and have been steadily evolving to attain a deeper understanding of cellular heterogeneity and transcriptional dynamics than the previously developed bulk RNA sequencing technologies. These advances have led to the discovery of novel cell types, developmental trajectories, and disease-associated states. In this review, we summarize the transformative potential and recent advancements in scRNA-seq approaches, especially focusing on those studies that integrate long-read sequencing platforms to simultaneously generate single-molecule and single-cell maps, providing more resolution to transcriptomic studies.

Advent of low throughput scRNA-seq technologies

The first protocol for scRNA-seq was developed in 2009, which relied on manually isolating individual cells, generating full-length cDNA from their mRNA, and subsequently sequencing it using short-read

Received: February 27, 2026. **Revised:** May 19, 2026. **Accepted:** May 21, 2026

sequencing technologies [1]. While it successfully demonstrated the possibility of sequencing the entirety of the mRNA in a single cell, the manual isolation protocol was only able to sequence a small quantity of cells at a time and preferentially amplified the 3′ end of the molecule due to the use of a single polymerase chain reaction (PCR) primer [2]. Single-cell Tagged Reverse Transcription (STRT) [3] was developed to increase the throughput of scRNA-seq by isolating single cells into a 96-well plate. Barcodes unique to each well were added to the 3′ end of the cDNA allowing 96 cells to be pooled and sequenced together on short-read sequencers. However, with the addition of the 3′ barcode on the cDNA needed for pooling, the full length of the transcript was no longer able to be sequenced because only the reads on the 3′ end containing the barcode can be related back to a single cell. Smart-Seq [2] was developed to address the inability of the previous methods to capture the full-length transcript utilizing the switching mechanism at 5′ end of RNA template (SMART) template switching technology to generate unbiased full-length cDNA of the mRNA. There were no barcodes used in Smart-Seq requiring each cell to be manually tracked, limiting throughput again to a few cells at a time. Smart-Seq2 [4] improved capture efficiency, PCR conditions, and reproducibility. Another protocol focused on reducing PCR bias, cell expression by linear amplification and sequencing (CEL-Seq) [5], had a unique development that allowed for in vitro transcription (IVT) to avoid performing PCR, skipping the potential for PCR bias. This was accomplished by performing single-cell reverse transcription that had a primer with an anchored polyT, unique barcode, the Illumina 5′ sequencing adapter, and T7 promoter. Pooling the cDNA from multiple samples together was then enough input material to successfully overcome the barrier of the quantity needed to perform IVT. This method was updated to CEL-Seq2 [6] where unique molecular indexes (UMIs) were added to ensure each mRNA was counted only once because the UMI could be used to identify all reads that were generated from the original template. Lastly, Smart-seq3 [7] then improved Smart-seq2 by including UMIs.

Increasing throughput with FACS and microfluidics

Due to the inability of short-read platforms to relate full-length transcripts to a single cell when multiplexing samples, subsequent scRNA-seq protocols focused on increasing throughput. Massively parallel RNA single-cell sequencing (MARS-Seq) [8] was designed to function in vivo and utilized fluorescence-activated cell sorting (FACS). FACS is a specialized flow cytometry process by which fluorescent antibodies bind to proteins on cells and the fluorescent signal read from a detector plate is used in conjunction with a deflector plate to sort cells. When this technology was applied, the throughput of cell isolation increased, filling a larger plate of 384 wells. After the integration of FACS into scRNA-seq, cell isolation through microfluidics became mainstream. Microfluidics allowed cells to be passed through a channel and then individually isolated within droplets that contain the necessary barcodes for sequencing. This automated hands-off technology further increased the feasible number of cells that could be isolated and prepared for sequencing. The droplet sequencing protocol (Drop-Seq) [9], which bound transcripts to beads within the droplet and performed sequencing preparation on the beads after the droplet was broken, indexing droplets (inDrop) [10], which performed sequencing preparation within the droplet, and the 10X Genomics platform [11] that increased cell capture efficiency and sensitivity performed sequencing preparation within the droplet, were popular applications of microfluidics to scRNA-seq.

Cell barcoding without individual cell isolation

Combinatorial barcoding, the newest development in short-read scRNA-seq, removes the need for the physical isolation of cells. The basis of combinatorial barcoding approaches is that when cells are repeatedly pooled and redistributed with unique barcodes introduced for each subset of cells during redistribution, these combinatorial barcodes have a low likelihood of being identical across multiple cells. Experimental complexity is then reduced while further increasing multiplexing with the dramatic increase in potential barcodes. The first combinatorial barcoding method to emerge was single-cell combinatorial indexing RNA sequencing (sci-RNA-seq) [12], where cells are fixed, permeabilized, and split across 96- or 384-well plates with FACS. Barcodes are added, and the cells undergo a round of pooling followed by redistribution via FACS into a second plate of wells to receive another barcode. The barcode combination can then be used to uniquely identify cells. It is possible for more than one cell to have the same barcode combination, and this likelihood can be adjusted by lowering the number of cells in each well during redistribution by FACS. The sci-RNA-seq protocol was later independently optimized where alternate reagents are used to increase efficiency [13]. A variant of sci-RNA-seq, sci-RNA-seq3 [14] incorporates a third round of barcoding, which also lowers the likelihood of multiple cells containing the same combination of barcodes while increasing the set of unique barcodes for even higher multiplexing capability. Split pool ligation-based transcriptome sequencing (SPLiT-seq) [15] is another protocol that uses three initial rounds of pooling and barcoding within wells. The barcoded cells are pooled once again and split as sublibraries and a final barcode is added. Sublibraries are used instead of a single library to split the large quantity of cells from different samples into smaller portions, reducing complexity while allowing each sublibrary to be handled individually for optimizations such as normalization. Additionally, Tiny-Sci [13] was developed as an adaptation to sci-RNA-seq specifically designed for working with small samples. These advancements enable researchers to sequence large volumes of cells in a highly multiplexed fashion, but due to the necessity of cellular barcodes and the use of short-read sequencing, the entire transcript could not be sequenced. Figure 1 provides a timeline of these technological developments.

Applying high-throughput long-read sequencing technologies for single-cell transcriptomics

Hybrid long-read/short-read approaches

Recent advancements in sequencing protocols successfully expanded the throughput of scRNA-seq, but these protocols were still limited in their ability to sequence the entire transcript. Instead of only altering the protocol, researchers began to utilize long-read sequencers with adaptations to the already developed protocols for short-read scRNA-seq. The first long-read single-cell protocol, single-cell isoform RNA sequencing (ScISOr-Seq) [16], utilized PacBio sequencing on the 10X Genomics platform to capture individual cells and append barcodes and UMIs. The cDNA library was then sequenced in a hybrid approach with some cDNA sequenced using Illumina, and the remainder was

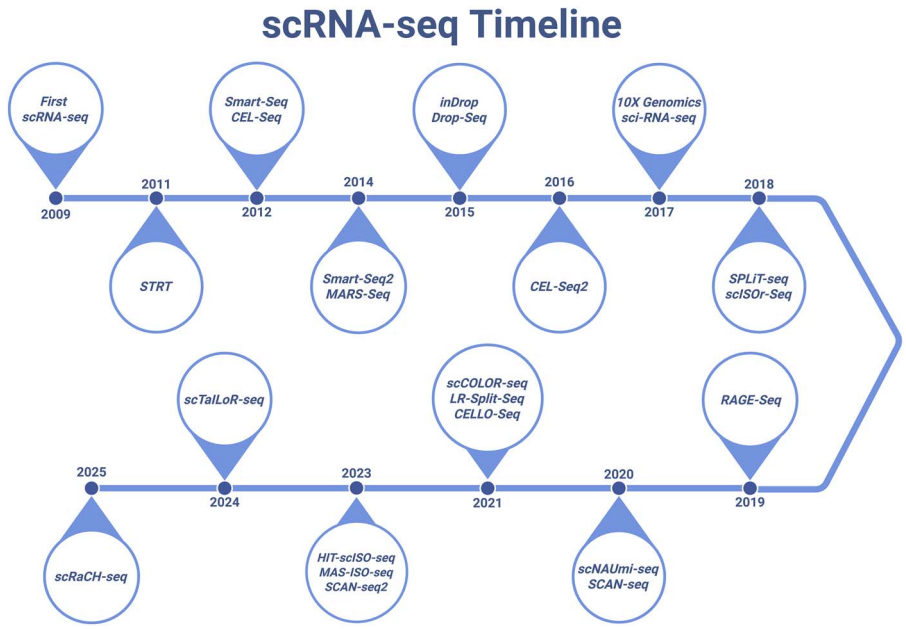

Figure 1 Timeline showing the advancement of scRNA-seq approaches to develop single molecule scRNA-seq techniques during the past decade. Created in BioRender. Schaeper, D. (2026) https://BioRender.com/s5mn13e.

sequenced using PacBio single molecule, real-time (SMRT) cells. The Illumina reads were used for clustering and expression profiles where the long-read data was used solely for isoform-level information. Repertoire and gene expression by sequencing (RAGE-Seq) [17] utilizes Oxford Nanopore Technologies (ONT) sequencing paired with short-read sequencing. Similar to ScISOr-Seq, a hybrid approach to sequencing was used, except the cDNA portion utilized ONT sequencing and hybrid capture to selectively enrich transcripts. The analysis is also similar where the short-read data is used for expression profiling, and the long-read data is analyzed for isoform and splicing information. These protocols either did not use UMIs [17] or did not perform UMI correction [16] for the long reads, so there was uncertainty about whether biases introduced by PCR were correctly handled. Single-cell nanopore sequencing with UMIs (ScNaUmi-seq) [18] was designed to use matched UMIs from Illumina sequencing looking only at those reads that were assigned to the same cell and aligned to the same genomic region from both sequencing datasets that yielded a cell barcode and UMI assignment accuracy of 99.3% compared to the Illumina data.

Standalone long-read scRNA-seq

The lower accuracy of third-generation sequencing compared to second-generation sequencing compelled researchers to use Illumina sequencing in conjunction with long-read sequencing to ensure proper barcode and UMI assignment, making it potentially cost-prohibitive. The first method developed to use standalone nanopore sequencing was single-cell corrected long-read sequencing (scCOLOR-Seq) [19]. Through the use of dimeric barcode and UMI oligos, where every nucleotide in the barcode was duplicated, sequencing errors within the barcodes were easily detected and corrected. The barcodes were also designed to have a large Hamming distance so that if there were multiple errors within the barcode, they could still be unambiguously identified. Using this method, barcode recovery was found to be over 80%. Another standalone method, long-read split pool ligation-based transcriptome sequencing (LR-Split-Seq) [20], was built upon the SPLiT-seq [15] protocol using combinatorial barcoding instead of the 10X Genomics platform to isolate and barcode cells. They demonstrated that it is as effective as short-read SPLiT-seq for detecting cell clusters despite the slightly lower accuracy of PacBio than Illumina. In addition to the reduced cost and ease of workflow, switching from a hybrid short and long-read sequencing approach to only long read was shown to increase recall when investigating transposable elements (TEs). A library of mouse embryonic stem cells sequenced by ScNaUMI-seq was unable to correctly map the short reads of TEs due to their repetitive nature resulting in no shared barcodes within these regions for long-read data analysis to build upon. Single-cell long RNA sequencing (CELLO-seq) [21] was developed as a standalone method to address this. It affixed 22 nt long UMI sequences with a repetitive pattern of RYN onto the transcripts, which allowed for better identification and error correction while preventing homopolymers, a low-accuracy sequence structure in ONT sequencing. Similar to scCOLOR-seq, these nanopore-specific adjustments to the UMIs alleviated the accuracy issues that can arise and enabled it to be a standalone method. An approach to generate higher accuracy nanopore reads became available with rolling circle to concatemeric consensus (R2C2) sequencing, a nanopore sequencing method that circularizes cDNA and produces consensus reads for improved accuracy, as an alternative to relying upon cleverly designed UMI sequences. The first study to utilize R2C2 with long-read scRNA-seq utilized standard 10X Genomics cell isolation [22]. They were able to assign 81% of reads to cells based on the barcodes, higher than even the Illumina-guided assignment rate of ∼67%. Of the 10×-UMIs in the validation Illumina dataset, 67% were in the ONT set for each cell, with a significantly higher read accuracy for the reads that did contain the same UMIs as the Illumina set. The cell clusters generated are also correlated with the clusters generated by the short-read sequencing. These studies provide evidence supporting standalone long-read scRNA-seq, thereby eliminating the need for matched short-read data.

Despite standalone long-read scRNA protocols' experimental validity, they still require larger amounts of input than short read and do not have the deep depth provided with Illumina short-read platforms because of both long-read platform limitations and the generation of non-barcoded reads. The higher amount of input needed was addressed by the single-cell amplification of full-length RNA nanopore sequencing (SCAN-seq) [23] protocol. It uses a pool of 48 cells labeled with nanopore-compatible barcodes that increased the input volume enough to satisfy the needs of long-read sequencing. SCAN-seq2 [24] was an improvement to SCAN-seq developed to increase sensitivity and throughput. Throughput was specifically addressed by high-throughput single-cell isoform sequencing (HIT-scISOseq) [25] designed for PacBio that integrates two extra steps to provide higher throughput. First, template switching oligo (TSO) artifacts are removed by PCR-based biotin-assisted capture. Second, multiple cDNA amplicons are concatenated into long SMRTbell inserts to fully utilize the long capacity of PacBio's circular consensus sequencing to sequence multiple transcripts in a single read. This was expanded in multiplexed arrays isoform sequencing (MAS-ISO-Seq) [26] and sequenced approximately 40 million full-length transcripts per SMRT Cell 8M flow cell.

Targeted long-read scRNA-seq

There are also standalone methods that perform targeted LR-scRNA-seq that are valuable when researchers need high coverage for specific RNA targets. The first of these is single-cell targeted isoform long-read sequencing (scTaILoR-seq) [27]. This method utilizes commercialized gene panels to enrich for genes of interest after 10X Genomics isolation and barcoding. This is followed by an artifact mitigation step reducing TSO–TSO byproducts using biotinylated PCR primers complementary to Read1 enabling pull-down and amplification of complete cDNA. These are sequenced using ONT sequencing and result in ∼95% of usable transcript reads mapped to the targeted genes. A different targeted method called single-cell rapid capture hybridization sequencing (scRaCH-seq) [28] has a unique focus on being able to utilize pre-indexed and stored cDNA from any 10X Genomics kit. Biotinylated probes bind to genes of interest and are captured with Streptavidin bead pull down to be further amplified. Only these targeted genes are passed on to sequencing with ONT. This protocol, however, does call for paired Illumina short-read sequencing to demultiplex and perform clustering because the targeted genes of interest are not sufficient on their own. The validation of their chosen enrichment genes yielded 92.6% read retention. It is shown that there is currently a diverse set of protocols that yield both whole transcriptome and targeted scRNA-seq with long-read sequencing platforms that provide an option of paired Illumina short-read data or function as standalone long-read data. Figure 2 displays the overarching concepts of these sequencing protocols.

Computational frameworks for analyzing single-molecule scRNA-seq

With the advent of long-read scRNA-seq, and different protocols for generating it, new analysis methods were developed to properly analyze the resulting data. The preprocessing stage of long-read single-cell RNA-seq analysis is centered on the accurate recovery of cellular barcodes and UMIs, which are critical for assigning reads to their cell of origin and determining accurate expression levels. The preprocessing steps directly impact the fidelity of downstream isoform discovery, variant detection, and expression analysis. A range of methods have recently been developed to address the challenges of barcode demultiplexing in the context of long-read data, particularly given the higher error rates associated with Oxford nanopore sequencing.

Early efforts focused on hybrid strategies that combine long-read data with matched short-read sequencing. Single-cell nanopore barcode (scNapBar) [29] exemplifies this approach by implementing a Bayesian framework that integrates read quality scores and barcodes prior to rescuing barcodes that might otherwise be missed. This hybrid strategy increases barcode recovery and ensures better representation of true cells, making it a valuable preprocessing step for studies that already generate complementary Illumina data. To optimize efficiency in hybrid workflows, scTagger [30] was introduced as a fast method for linking short-read and long-read barcodes in split-library designs. By using a trie-based data structure with error-tolerant matching, scTagger reduces the computational burden while maintaining high accuracy. This makes it feasible to process millions of reads and scale long-read single-cell studies to larger cohorts.

Recent tools have moved beyond hybrid reliance by enabling long-read-only barcode identification, thereby simplifying experimental design and reducing cost. Barcode identification from long reads for analyzing single-cell gene expression (BLAZE) [31] is a key development in this direction, accurately identifying 10X Genomics barcodes directly from Nanopore reads without the need for matched short-read data. Using adapter and polyT detection coupled with stringent quality filtering, BLAZE generates a high-confidence whitelist. This approach not only eliminates the dependency on Illumina sequencing but also minimizes false positives, ensuring that downstream clustering reflects biologically meaningful cell populations. A more generalizable solution is provided by Flexiplex [32], which is designed as a versatile demultiplexing and sequence search tool. Built on Levenshtein distance–based alignment, Flexiplex accommodates sequencing errors, supports discovery of unknown barcodes, and uniquely handles chimeric read splitting. Its flexibility makes it suitable not only for cellular barcode identification in long-read single-cell datasets but also for broader omics applications such as clonal tagging, UMI identification, and pooled sample demultiplexing.

Finally, Tranquillyzer [33] represents a new wave of neural network–driven approaches to preprocessing. By combining convolutional neural networks, bidirectional long short-term memory (BiLSTM) models, and conditional random fields, Tranquillyzer performs structural annotation of long-read transcriptomes, identifying barcodes, UMIs, and adapters with high accuracy. In benchmarking, it achieved over 99% annotation accuracy and more than 91% demultiplexing efficiency, while also detecting and filtering chimeric molecules. This shift from heuristic to machine learning–based models highlights the field's progression toward scalable, adaptive solutions capable of handling increasingly complex single-cell transcriptome datasets.

The discussed long-read single-cell RNA-seq preprocessing tools illustrate a continuum in the evolution of preprocessing methodologies for long-read single-cell RNA-seq. From hybrid barcode rescue methods to flexible error-tolerant demultiplexers and AI-driven structural annotation, they collectively establish the foundation upon which post-processing analyses such as isoform discovery, fusion detection, and variant calling can be reliably built.

Once reads have been accurately assigned to cells, post-processing steps extract the full biological value of long-read single-cell RNA-seq data. Some of these tools have previously been developed for bulk

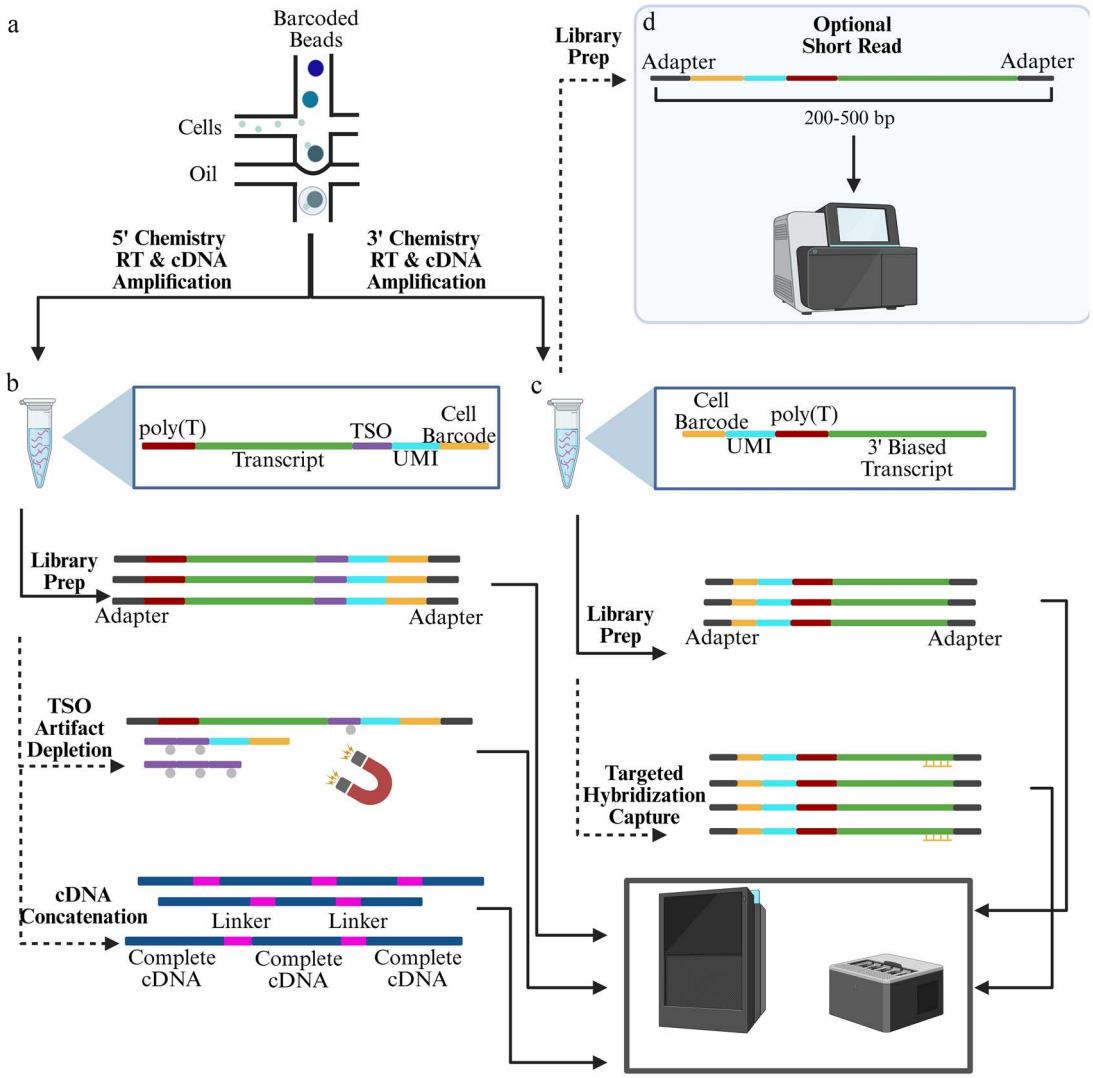

Figure 2 Schematic overview of the wet lab workflows based on the 10X Genomics platform. (a) 10X Genomics cell isolation. (b) Amplicon structure and optional modifications from developed protocols using the 5′ chemistry to be integrated with long-read sequencing of transcriptomes. (c) Amplicon structure and optional modifications from developed protocols using the 3′ chemistry to be integrated with long-read sequencing of transcriptomes. (d) Optional paired short read sequencing using the same library. Created in BioRender. Schaeper, D. (2026) https://BioRender.com/fj3lp1g.

RNA-seq from long-read sequencing and have been applied to long-read scRNA-seq, while others have been specifically developed for it. A major focus is isoform discovery and quantification, where long-read RNA-seq tools such as Mandalorion [34], an alignment-based tool, Isosceles [35], a splice graph with an expectation maximization quantification method, IsoQuant [36], which creates an intron graph that discovers transcripts through path traversal, Bambu [37], a machine learning approach to transcript discovery, and technology agnostic long-read analysis pipeline for transcriptomes (TALON) [38], a tool that corrects reads and performs annotation of the reads using an SQLite database seeded from the Encyclopedia of DNA Elements (GENCODE) that can be expanded with novel isoforms, can identify transcript structures at single-cell resolution. Single-cell omics for transcriptome characterization (SCOTCH) [39] is a long-read scRNA-seq specific tool that identifies present known and novel isoforms by aligning reads to isoforms and splitting the alignments to non-overlapping sub-exons. Unique combinations of these sub-exons define isoforms; thus, both known isoforms and those with novel splicing patterns can be identified. These methods leverage full-length reads to resolve alternative splicing, quantify isoform usage, and uncover transcript diversity that short-read approaches often miss. Another key application is gene fusion detection, particularly in cancer. Just another fusion finder algorithm for long-read (JAFFAL) [40] applies long-read–adapted alignment and exon-boundary anchoring to detect high-confidence fusions while filtering artifacts, enabling identification of clinically relevant events such as RUNX1-RUNX1T1 in leukemia, even at the single-cell level.

Beyond transcripts and fusions, variant detection links genotypes and phenotypes within the same cell. Long-read somatic variant caller (LongSom) [41] simultaneously identifies single-nucleotide variants, copy number changes, and fusions from long-read scRNA-seq, reconstructing clonal architecture and highlighting genetic heterogeneity across cells. Together, these post-processing approaches move analysis from cell-assigned reads to interpretable features: isoform atlases, fusion landscapes, and clonal genotypes. They establish the intermediate layer of the workflow, providing outputs

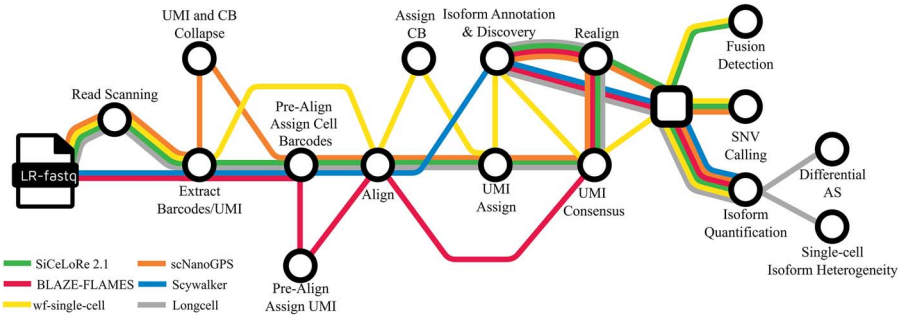

Figure 3 Comparison of the published end-to-end workflows for analyzing long-read single cell RNA-sequencing datasets. Each line represents a step-by-step progression through the tool's workflow detailing the various major computational steps. Open circles mark steps in the pipeline, labeled with the corresponding step.

that can be integrated into higher-level analyses and disease models.

Along with standalone tools, multiple end-to-end workflows have been developed to facilitate easy analysis of long-read scRNA-seq data. Longcell [42], single-cell nanopore genotype phenotype sequencing (scNanoGPS) [43], and workflow single cell (wf-single-cell) [44] have been designed to accept ONT data. Full-length analysis of mutations and splicing (FLAMES) [45], Scywalker [46], and single cell long read (SiCeLoRe) 2.1 [18] take either ONT or PacBio data as input. While FLAMES is a standalone workflow, it is common practice to demultiplex with BLAZE instead of having FLAMES perform its own demultiplexing. Scywalker goes beyond performing long-read analysis and performs cell clustering as well. While a toolkit and not a single workflow, single cell isoform analysis tools (scISA-Tools) [25] is a notable single resource that is designed to analyze long-read scRNA-seq data from PacBio sequencers. The end-to-end workflows are described in Fig. 3 along with the tools in Table 1.

Together, these workflows, pipelines, and toolkit demonstrate the field's transition toward ecosystem-level solutions. Whether through end-to-end automation, modular isoform splicing and quantification, or genotype–phenotype integration, they unify the strengths of pre-processing and post-processing into cohesive analytical strategies that can be tailored to diverse biological questions.

Applications of single-molecule scRNA-seq

Long-read scRNA-seq has been utilized across many fields of study to further expand the understanding of the field and has led to major clinical insights in recent years. Specifically, the long reads have provided additional resolution to elucidate the presence of different isoforms, alternative splicing, and gene fusions (Table 2). Along with structural information, the long reads provided complete mutational profiles of cells that could not be captured when sequencing only one tail of the transcript.

Cancer transcriptomics

Researching cancer transcriptomics is a large focus of long-read scRNA-seq because of the technique's ability to capture gene fusions and novel isoforms that can occur in the disease, which short-read sequencing struggles to detect. A study of colorectal cancer (CRC) revealed many insights into the isoforms present. Through long-read scRNA-seq they were able to identify isoforms associated with epithelial cell subtypes enterocyte, goblet, and BEST4 lineages, determined that each tumor contained a substantial proportion of three subtypes with varying differentiation and proliferation, highlighting intratumor heterogeneities [47]. A second study on CRC utilized the long reads to identify a trend of 3′-UTR shortening within cancer cells that was highly prevalent in stem/TA-like cells. These shortened 3′-UTRs were associated with RNA processing and metabolic processes that were correlated with increased RNA expression levels, demonstrating a potential purpose of this modification. They also discovered alternative splicing alterations with reduced intron retention, also within the stem/TA-like cells. Finally, they found isoform switches in oncogenic differential transcript usage (DTU) genes that frequently altered coding sequences and protein domains that could influence cancer progression [49]. A third study focused on genomic alterations in ovarian cancer and their effects on the coding potential of the transcripts. Firstly, they identified a gene fusion, IGF2BP2::TESPA1, that was misidentified in short-read data. While investigating isoforms of coding mRNA from long-read scRNA-seq, numerous isoforms of protein-coding genes were found to be wrongly accounted for as protein-coding, and instead potentially serve as microRNA (miRNA) sponges and competing endogenous RNAs (ceRNAs). In addition, 49% of the genes exhibiting a significant isoform switch between cancer and healthy cells transitioned from protein coding to a noncoding isoform demonstrating large functional changes that can be misrepresented in short-read scRNA-seq [48]. Long-read scRNA-seq was leveraged to study subclones in chronic lymphocytic leukemia (CLL) in a fourth study using the improved coverage of somatic mutations. The BTK^{C481S} mutation is a common cause of resistance to treatment utilizing a BTK inhibitor that can occur in relapse patients. This and additional CLL driver mutations can be missed while performing short-read scRNA-seq, and the researchers were able to use the full-length transcript sequence to not only investigate this mutation but further characterize the subclone structure within patients with the ability to identify potentially resistant subclones. This refined a previous study that used a combination of whole-exome sequencing and short-read scRNA-seq [50]. As a result of the increased resolution provided by long-read scRNA-seq, multiple research groups were able to further capture previously unidentifiable significant alterations within tumor cells.

Neuroscience

The brain is another complex tissue that is studied to understand the mechanisms behind neurological diseases. Using long-read scRNA-seq, researchers were able to further expand upon the available knowledge of transcriptomics in the brain. One study aimed to

Table 1 A guide to computational tools for single-cell isoform sequencing with long reads.

Tool	Long-read sequencing platform	Usable sequencing protocols	Core functionality	Benchmarking/validation	Strengths	Limitations	Availability/implementation
Flexiplex [31]	ONT, PacBio	Any	Demultiplexing, UMI identification	Benchmarked versus scTagger, FLAMES, BLAZE; simulated + real scRNA-seq data; processed 25 M ONT reads in 2.6 h with <1 GB RAM	Versatile, lightweight, customizable; supports noisy ONT	Slower than scTagger on huge data; relies on barcode flanks	https://davidsongroup.github.io/flexiplex/
BLAZE [30]	ONT	10X Genomics	Demultiplexing	Validated on 61 M ONT reads from hiPSC	Works without short reads; efficient	Accuracy drops with extreme ONT errors; requires barcode whitelist	https://github.com/shimlab/BLAZE
ScTagger [29]	ONT, PacBio	10X Genomics, Illumina paired data needed	Demultiplexing	Real + simulated datasets (testis biopsies, infertility patients); outperformed FLAMES and brute-force; accurate with edit distance ≤2	Fast and accurate; scalable to thousands of cells	Specific to 10×; memory heavy on very large datasets	https://github.com/vpc-ccg/scTagger/
Tranquillyzer [32]	ONT, PacBio	Any	Demultiplexing	Simulated datasets (up to 100 M reads, 2.5 kb length); >99.7% annotation accuracy, >91% demultiplex efficiency; benchmarked versus SiCeLoRe, scNanoGPS, wf-single-cell	High accuracy; robust to structural noise; GPU-accelerated	Needs training per library type; GPU memory intensive; preprint stage	https://github.com/huishenlab/tranquillyzer.git
ScNapBar [28]	ONT	10X Genomics, Illumina paired data needed	Demultiplexing, UMI identification	Benchmarked on simulated and experimental Nanopore + Illumina data; improved barcode recovery versus naïve approaches	Recovers more true barcodes, reduces false positives	Needs short-reads for accurate demultiplexing; not standalone for ONT-only	https://github.com/dieterich-lab/single-cell-nanopore
SCOTCH [38]	ONT, PacBio	10X Genomics, Parse Biosciences	Isoform mapping, isoform detection, isoform quantification, DTU	Simulations + PBMC and organoid datasets; outperforms IsoQuant and FLAMES; detects DTU/isoform switches	High isoform mapping accuracy; robust across library types	Overestimates novel isoforms; still in preprint	https://github.com/WGLab/SCOTCH
Mandalorion [33]	ONT, PacBio	Any R2C2 or ISO-Seq	Isoform detection, isoform quantification, isoform annotation, TSS association	Simulated data + UHR Iso-Seq; recall/precision >90%; TPM r≈0.992	High recall and precision; annotation-independent	Slight undercounting; terminal exon detection issues	https://github.com/christopher-vollmers/Mandalorion
Isosceles [34]	ONT, PacBio	Any long-read protocol	Isoform/Transcript quantification and discovery	Simulated/synthetic benchmark with spike-ins; outperforms Bambu and IsoQuant	High accuracy across scales; efficient	Reduced performance on very low expression	https://github.com/Genentech/Isosceles
LongSom [40]	ONT, PacBio	Any	Variant detection	Applied to ovarian cancer scRNA-seq; validated with scWGS and matched DNA	Detects multiple variant types simultaneously; reconstructs clonal heterogeneity	Requires high-quality LR scRNA-seq; computationally heavy	https://github.com/cbg-ethz/LongSom

(*Continued*)

Table 1 Continued

Tool	Long-read sequencing platform	Usable sequencing protocols	Core functionality	Benchmarking/validation	Strengths	Limitations	Availability/implementation
JAFFAL [39]	ONT, PacBio	Any	Gene Fusion	Benchmarked on cancer cell lines, and it was compared against other long-read and short-read tools	Accurate detection of known and novel fusions	May miss low expression fusions; less optimized for single-cell	https://github.com/Oshlack/JAFFA
FLAMES [44]	ONT, PacBio	Protocols with barcode whitelist	Demultiplexing, alignment, UMI assignment and consensus, isoform quantification, isoform discovery, isoform annotation, SNV calling	Outperforms TALON/FLAIR/StringTie2; high UMI correlation with Illumina	Multi-omics integration; novel isoform detection	Lower throughput than Illumina; false positives possible	https://github.com/LuyiTian/FLAMES
Longcell [41]	ONT	10X Genomics Proved, may be compatible with others	Barcode and UMI extraction, demultiplexing, UMI consensus, isoform quantification	Simulation, PacBio–Nanopore reference, outperforms IsoQuant and Sicelore2	Accurate quantification at spatial resolution; robust to ONT errors	Dependent on ONT accuracy; limited discovery	https://github.com/yuntianf/Longcell
ScNanoGPS [42]	ONT	10X Genomics	QC, extract barcodes and UMIs, make barcode list, demultiplex, align, UMI consensus, isoform annotation, isoform quantification, SNV	Applied to human tumor samples; validated same-cell mutations + expression	Enables genotype–phenotype coupling at single-cell resolution	ONT error profile can complicate variant calls	https://github.com/gaolabtools/scNanoGPS
Scywalker [45]	ONT, PacBio	10X Genomics	Demultiplexing, alignment, isoform detection, isoform quantification, cell specific QC, cell type annotation, pseudobulk cell type counts	Tested on human brain and Arabidopsis datasets; benchmarked versus existing ONT workflows; shows correlation with Illumina SR	Scalable (>10 k cells); full workflow; integrates QC and reporting	Still new; dependent on IsoQuant for isoform detection	github.com/derijkp/scywalker
SiCeLoRe 2.1 [18]	ONT	10X Genomics	Barcode assignment, alignment, UMI assignment, UMI consensus, fusion detection, isoform quantification, isoform discovery	E18 mouse brain, correlated well with Illumina SR	Updated to short-read free analysis, Nextflow implementation	Needs high-quality data for barcodes	https://github.com/ucagenomix/sicelore
wf-single-cell [43]	ONT	10X Genomics	QC, extract barcodes and UMIs, align, demultiplex, isoform discovery, UMI assignment, UMI consensus, fusion detection, SNV calling, isoform quantification	Utilized BLAZE benchmarking	Simple Nextflow implementation	Only viable for ONT	https://github.com/epi2me-labs/wf-single-cell
scISA-Tools [24]	PacBio	HIT-scISOseq, ScISOr-Seq	QC, transcript identification, barcode and UMI extraction and correction, isoform clustering, isoform quantification	≥10 M reads per SMRT Cell, applied to 3375 corneal limbus cells	Specific design leads to high accuracy	Requires external tools meshed within toolkit	https://github.com/shizhuoxing/scISA-Tools

Table 2 Key applications in long-read single-cell RNA sequencing.

Study	Research area	Experimental design basis	Long-read sample details	Coupled with short read	Species	Novel isoforms/key discoveries	Clinical relevance	Data availability
[46]	Cancer	PacBio Iso-Seq	4 CRC samples, 3 normal	Yes	Human	125 205 unique isoforms; 394 dysregulated transcript structures in tumor epithelium; isoform-specific RNA editing; 330 isoforms supported by MS; 22 shared neoepitopes proposed	Neoepitope candidates (vaccine design); isoform-level biomarkers; editing signatures	PRJEB68074
[47]	Cancer	SCAN-seq, SCAN-seq2	12 primary CRC tumors and matched normal controls	No	Human	Increased transcript complexity in cancer; 3-UTR shortening; decreased intron retention; subtype (iCMS3) with higher splicing-factor activity; allele-specific expression with functional validation	Splicing-program differences suggest therapeutic targets; allele-specific regulation in cancer genes.	HRA006041
[48]	Cancer	MAS-Seq	Pretreatment and released samples from 6 CLL patients with acquired BTKC481S	No	Human	Long-read enables cell-level genotyping of subclone-defining mutations and mapping of subclone-specific programs; additional drivers co-occur	Supports tracking resistance evolution; subclone-specific expression could guide therapy	GSE259253
[49]	Cancer	HIT-scISOseq with cDNA concatenation and biotin PCR	3 HGSOC omental metastases 2 from distal tumor free tissue	Yes	Human	∼152 k isoforms; >52 k novel; cell-type–specific isoform/APA	Enables isoform-level biomarkers; fusion/mutation detection at single-cell resolution.	EGAS00001006807
[50]	Neurology	ScISOrSeq	2 hippocampus and 2 prefrontal cortex	Yes	Mouse	Hundreds of genes with differential isoform expression across regions/cell types; distinct transcript start site usage in choroid plexus epithelium	Foundational atlas for brain region– and cell type–specific isoform regulation; resource for neurodevelopmental/neurological disease hypotheses.	GSE158450
[51]	Neurology	scIso-seq	7 Cerebral organoid samples at different time points	Yes	Human	∼31 000 uncatalogued/novel isoforms; 4531 cell-type-specific splicing/exon events; autism-associated exons enriched for de novo mutations	Links isoform/exon usage to ASD risk; insights into early neurodevelopmental regulation	PRJNA947248
[52]	Developmental Biology	SCAN-seq HIT-scISOseq hybrid approach	4 batches of blastomeres	Yes (Barcodes with Sanger)	Mouse	3′ partial transcripts (stop-codon-less) enriched in oocyte/zygote; 3894 TE loci with dynamic expression	Insights into MZT and TE regulation in early embryos; mechanistic foundation for infertility/embryology studies.	GSE250381

(*Continued*)

Table 2 Continued

Study	Research area	Experimental design basis	Long-read sample details	Coupled with short read	Species	Novel isoforms/key discoveries	Clinical relevance	Data availability
[53]	Developmental Biology	ScISOrSeq	3 OA and 3 NOA samples	Yes	Human	Detected 130 426 (OA) and 49 508 (NOA) full-length transcripts; extensive novel transcript and novel gene discovery; mapped full-length isoforms across germ/somatic lineages	Improves molecular stratification of male infertility by revealing isoform-level differences across spermatogenic stages and cell types.	No accession supplied
[54]	Infectious Diseases	10X Genomics plus additional amplification PCR with PacBio sequencing	4 samples	Yes	*P. vivax* from *Saimiri boliviensis*, Anopheles stephensi, and *Anopheles freeborni*	Widespread isoform diversity via UTR length variation and splicing; stage-specific isoform expression; examples include unannotated introns and alternative TSS altering protein predictions.	Refines stage-specific transcript models; supports vaccine/drug-target discovery pipelines in vivax malaria.	PRJNA863611
[55]	Diagnostic Utility	MAS-ISO-seq with cDNA post-TSO artefact removal enriched for PIK3CA	1 sample	Yes	Human	Long reads captured the H1047Y locus in 15.3% of barcoded cells versus 1.2% with short reads; variant expression enriched in PAX3+ fibroblast-like and undifferentiated keratinocyte populations.	Directly identifies a targetable PIK3CA variant and its expressing cell states; informs therapy (e.g. alpelisib/mTOR inhibitors) and demonstrates diagnostic utility of targeted single-cell long-read genotyping.	GSE253153

generate an isoform atlas of the postnatal mouse brain. Applying long-read scRNA-seq to the hippocampus and prefrontal cortex, the researchers identified differentially expressed isoforms across 45 different cell types. Additionally, they discovered that while cell type generally determines microexon presence, the brain region can override this as demonstrated by Nsfl1c where it is exhibited not only in neuronal cells but widely throughout the hippocampal region [51]. Another study interrogated human cerebral organoids for cell-type-specific and autism related exons where the full-length transcriptome data expanded the isoform catalog for cell types. A total of 4820 exons that were unannotated in GENCODE v.40 and RefSeq were identified. A total of 3628 of these exons were discovered within known genes, 111 of which were in Simons Foundation Autism Research Initiative Autism Genes. In addition to these, 1527 unannotated splice isoforms were also found in autism-associated genes. Eight unannotated isoforms were present in PTEN, a gene where mutations are correlated with autism [52]. These findings suggest that many previously unannotated exons and isoforms in the neuronal context can be critical components of protein expression regulation with potential clinical significance.

Developmental biology

The early stages of development were studied by sequencing blastomeres from the mouse oocyte to blastocyst stage. Long-read scRNA-seq data of these cells were studied to examine isoform switching and TE expression during development. Using long-read scRNA-seq, isoform switch events were discovered to predominantly occur transitioning between the zygote to early 2-cell (E2C), and the E2C to late 2-cell (L2C) stage during major and minor zygotic genome activation. The genes that underwent isoform switching events were enriched for the gene ontology terms cytoplasmic translation, cell division, mRNA and DNA metabolic processes among others. When investigating isoforms, they noticed that as the embryo develops, the isoform diversity present tends to decrease. Interestingly, there was a large abundance of 3′ partial transcripts in the oocytes and zygotes. These genes with partial transcripts were enriched for functions crucial to embryo development. TE expression was analyzed for stage-specific expression patterns, and 3894 TE loci were classified into five clusters: higher expression in oocytes and zygotes, predominantly in E2C stage, high expression in L2C to 4-cell stage, high expression in 8-cell to blastocyst stage, and mouse embryonic stem cell-specific TEs [53]. A second study focused on development in the context of spermatogenesis in obstructive and non-obstructive azoospermia (OA, NOA), and Sertoli cell-only (SCOS). They identified >100 000 novel transcripts and > 36 000 novel genes in OA allowing for an increased understanding of the molecular mechanism of spermatogenesis. 22 transcripts were identified to be predominantly expressed in the pachytene stage. At the isoform level, SCOS-specific isoforms were confirmed to be present [56]. Researchers investigating isoform expression through long-read scRNA-seq in hematopoietic stem and progenitor cells found that over half of the genes in the hematopoietic hierarchy, as well as transcription factors in established networks, have multiple isoforms. These isoforms create additional complexity when building gene regulatory networks, especially in the hematopoietic stem cells. Complexes such as Lmo2/Gata2/Ldb1/Tal1 may be subjected to regulation with these isoforms, and this added complexity must be accounted for when building regulatory networks to ensure accuracy. [54]. Together, the transcriptome of cells involved in developmental processes has been further characterized in both a normal state providing novel insights and reference data and in a disease condition allowing for more clinical insights.

Infectious diseases

There has also been an application of long-read scRNA-seq to a single-celled organism with a complex life cycle. *Plasmodium vivax* infections are heterogenous populations of the parasite across different developmental stages that each have distinct transcriptional profiles, and long-read scRNA-seq provided the proper resolution to separate and characterize these different stages. Nearly 10,000 transcripts were identified from the blood stage parasite and ∼800 from sporozoites, where some of these transcripts suggest additional unannotated isoforms of known genes. Additionally, previous studies have been unable to fully capture the 3′ and 5′ UTR due to limitations of short-read sequencing, but this study was able to successfully capture these regions. Finally, stage-specific isoforms were found to be present across their data [55]. This shows that long-read scRNA-seq has a wide range of applications, here demonstrating that it can be used to expand the knowledge of difficult to characterize organisms.

Precision medicine

Finally, long-read scRNA-seq technology was shown to be valuable in a precision medicine context. Low-frequency mosaicism has the potential to make PIK3CA-related overgrowth spectrum (PROS) disorders difficult to diagnose, and long-read scRNA-seq was applied for a patient with suspected PROS but was receiving negative diagnostic tests. The PIK3CA mutation p.His1047Tyr was eventually found with a variant allele frequency of 11.9%. This mutation was found to be enriched in fibroblast-like cells characterized by neural crest marker expression. Interestingly, the clusters of cells expressing this mutation were enriched for cells within the G1 phase of the cell cycle [57]. This demonstrated that long-read scRNA-seq can be utilized at the individual level to diagnose disease.

Challenges and future prospects for high resolution single molecule transcriptomics

While there has been continuous improvement in the field of long-read scRNA-seq, there are still some challenges that could be addressed in both the wet and dry lab. The long-read platforms are continually improving; however, they still have lower throughput compared to short-read sequencing, which leads to less coverage of the genes being sequenced, thereby resulting in limited clinical insights. Also, the lower accuracy of the long-read sequencing platforms means low-frequency variants have a higher likelihood of being missed with long-read versus short-read sequencers. The ability to recall barcodes is negatively impacted by the lower read accuracy as well, meaning not all the sequenced data will be recoverable in the case of barcodes containing multiple errors. Finally, in the current form of typical long-read scRNA-seq protocols, RNA is converted to cDNA before sequencing, where information related to chemical modifications on RNA is completely erased. However, direct RNA sequencing (DRS) platforms such as those from ONT have the ability to directly sequence RNA and map modifications but have been underutilized due to challenges associated with cost and the amount of input material required for performing DRS. To allow this additional information to be accessible, there is a need to further develop methods and protocols that can

optimize the amount of input RNA required for direct RNA sequencing from single cells.

While there have been tools developed to identify cellular barcodes that perform well, accurately identifying the UMI has been less rigorously investigated. This is because the cell barcodes are generated from a known list whereas UMIs are generated with random bases thereby adding significant complexity. There have been adaptations to the experimental approaches to overcome the limitation of using standard barcoding approaches, but due to the popularity and ease of adoption of the 10X Genomics system, most of the computational methods have been tailored for these datasets resulting in a need for novel methods which can be applied to a wide variety of long-read scRNA-seq datasets. There is also a need for comprehensive third-party benchmarking to identify the ideal combination of the tools developed for analyzing long-read scRNA-seq so that the maximum value of the data is reached.

In the future, these current methods can be used to further elucidate biological mechanisms as well as the underlying dysregulation in diseases, with the increased range, completeness, and type of RNA that are captured in long-read scRNA-seq assays. Implementation of direct long-read scRNA-seq in the future will further enable the study of fine-grained regulatory functions comprehensively through SNPs, isoform usage, and RNA modifications within a single experiment largely improving functional understanding.

Key Points

- Current challenges of short-read approaches have resulted in the development of long-read protocols which can sequence full-length transcripts at a high resolution while maintaining the high throughput and multiplexing advancements of short-read scRNA-seq.
- Computational methods and frameworks were developed in parallel to analyze the newly generated long-read scRNA-seq data.
- The utility of the technology is demonstrated through applications in cancer transcriptomics, neuroscience, developmental biology, infectious diseases, and precision medicine.

Acknowledgements

We would like to thank members of the Jangalab for providing initial peer review. Additionally, we would like to thank Dr. Nicole Martinez and Nicolas Robalin for their feedback. Generative AI was used to check the grammar in some sections.

Author contributions

David Schaeper (Literature review, Writing—original draft & editing), Upol Chowdhury (Literature review, Writing—original draft, Writing—review & editing), and Sarath Chandra Janga (Conceptualization, Writing—review & editing)

Funding

None declared.

Data availability

Data referenced in this article are made available from their respective studies and are linked within Table 2.

References

1. Tang F, Barbacioru C, Wang Y *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009;**6**:377–82. https://doi.org/10.1038/nmeth.1315
2. Ramskold D, Luo S, Wang Y-C *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* 2012;**30**:777–82. https://doi.org/10.1038/nbt.2282
3. Islam S, Kjällquist U, Moliner A *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* 2011;**21**:1160–7. https://doi.org/10.1101/gr.110882.110
4. 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
5. Hashimshony T, Wagner F, Sher N *et al.* CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* 2012;**2**:666–73. https://doi.org/10.1016/j.celrep.2012.08.003
6. Hashimshony T, Senderovich N, Avital G *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol* 2016;**17**:77. https://doi.org/10.1186/s13059-016-0938-8
7. Hagemann-Jensen M, Ziegenhain C, Chen P *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat Biotechnol* 2020;**38**:708–14. https://doi.org/10.1038/s41587-020-0497-0
8. Jaitin DA, Kenigsberg E, Keren-Shaul H *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 2014;**343**:776–9. https://doi.org/10.1126/science.1247651
9. Macosko EZ, Basu A, Satija R *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015;**161**:1202–14. https://doi.org/10.1016/j.cell.2015.05.002
10. Klein AM, Mazutis L, Akartuna I *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 2015;**161**:1187–201. https://doi.org/10.1016/j.cell.2015.04.044
11. Zheng GX, Terry JM, Belgrader P *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;**8**:14049. https://doi.org/10.1038/ncomms14049
12. Cao J, Packer JS, Ramani V *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 2017;**357**:661–7. https://doi.org/10.1126/science.aam8940
13. Martin BK, Qiu C, Nichols E *et al.* Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nat Protoc* 2023;**18**:188–207. https://doi.org/10.1038/s41596-022-00752-0
14. Cao J, Spielmann M, Qiu X *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 2019;**566**:496–502. https://doi.org/10.1038/s41586-019-0969-x
15. Rosenberg AB, Roco CM, Muscat RA *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 2018;**360**:176–82. https://doi.org/10.1126/science.aam8999
16. 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
17. 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. https://doi.org/10.1038/s41467-019-11049-4

18. Lebrigand K, Magnone V, Barbry P *et al.* High throughput error corrected Nanopore single cell transcriptome sequencing. *Nat Commun* 2020;**11**:4025. https://doi.org/10.1038/s41467-020-17800-6
19. Philpott M, Oppermann U, Cribbs AP. Long-read single-cell sequencing using scCOLOR-seq. *Methods Mol Biol* 2023;**2632**:259–67. https://doi.org/10.1007/978-1-0716-2996-3_18
20. 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. https://doi.org/10.1186/s13059-021-02505-w
21. Berrens RV, Yang A, Laumer CE *et al.* Locus-specific expression of transposable elements in single cells with CELLO-seq. *Nat Biotechnol* 2022;**40**:546–54. https://doi.org/10.1038/s41587-021-01093-1
22. Volden R, Vollmers C. Single-cell isoform analysis in human immune cells. *Genome Biol* 2022;**23**:47. https://doi.org/10.1186/s13059-022-02615-z
23. Fan X, Tang D, Liao Y *et al.* Single-cell RNA-seq analysis of mouse preimplantation embryos by third-generation sequencing. *PLoS Biol* 2020;**18**:e3001017. https://doi.org/10.1371/journal.pbio.3001017
24. 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. https://doi.org/10.1038/s41421-022-00500-4
25. Shi ZX, Chen ZC, Zhong JY *et al.* High-throughput and high-accuracy single-cell RNA isoform analysis using PacBio circular consensus sequencing. *Nat Commun* 2023;**14**:2631. https://doi.org/10.1038/s41467-023-38324-9
26. 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
27. 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. https://doi.org/10.1038/s41467-024-51252-6
28. 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
29. Wang Q, Bönigk S, Böhm V *et al.* Single-cell transcriptome sequencing on the Nanopore platform with ScNapBar. *RNA* 2021;**27**:763–70. https://doi.org/10.1261/rna.078154.120
30. 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
31. 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
32. Cheng O, Ling MH, Wang C *et al.* Flexiplex: a versatile demultiplexer and search tool for omics data. *Bioinformatics* 2024;**40**:btae102. https://doi.org/10.1093/bioinformatics/btae102
33. Semwal A, Morrison J, Beddows PI *et al.* Tranquillyzer: a flexible neural network framework for structural annotation and demultiplexing of long-read transcriptomes. bioRxiv. 2025.
34. Volden R, Schimke KD, Byrne A *et al.* Identifying and quantifying isoforms from accurate full-length transcriptome sequencing reads with Mandalorion. *Genome Biol* 2023;**24**:167. https://doi.org/10.1186/s13059-023-02999-6
35. Kabza M, Ritter A, Byrne A *et al.* Accurate long-read transcript discovery and quantification at single-cell, pseudo-bulk and bulk resolution with isosceles. *Nat Commun* 2024;**15**:7316. https://doi.org/10.1038/s41467-024-51584-3
36. 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
37. 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
38. Wyman D, Balderrama-Gutierrez G, Reese F *et al.* A technology-agnostic long-read analysis pipeline for transcriptome discovery and quantification. bioRxiv. 2020;672931.
39. Xu Z, Qu H-Q, Chan J *et al.* Single-cell omics for transcriptome characterization (SCOTCH): isoform-level characterization of gene expression through long-read single-cell RNA sequencing. bioRxiv. 2025.
40. Davidson NM, Chen Y, Sadras T *et al.* JAFFAL: detecting fusion genes with long-read transcriptome sequencing. *Genome Biol* 2022;**23**:10. https://doi.org/10.1186/s13059-021-02588-5
41. 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
42. 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
43. Shiau CK, 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. https://doi.org/10.1038/s41467-023-39813-7
44. Labs EM. *Wf-Single-Cell*2025.
45. 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. https://doi.org/10.1186/s13059-021-02525-6
46. De Rijk P, Watzeels T, Küçükali F *et al.* Scywalker: scalable end-to-end data analysis workflow for long-read single-cell transcriptome sequencing. *Bioinformatics* 2024;**40**:btae549. https://doi.org/10.1093/bioinformatics/btae549
47. 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. https://doi.org/10.1016/j.xgen.2024.100641
48. 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. https://doi.org/10.1038/s41467-023-43387-9
49. Lu P, Zhang Y, Cui Y *et al.* Systematic characterization of full-length RNA isoforms in human colorectal cancer at single-cell resolution. *Protein Cell* 2025;**16**:873–95. https://doi.org/10.1093/procel/pwaf049
50. Black GS, Huang X, Qiao Y *et al.* Long-read single-cell RNA sequencing enables the study of cancer subclone-specific genotypes and phenotypes in chronic lymphocytic leukemia. *Genome Res* 2025;**35**:686–97. https://doi.org/10.1101/gr.279049.124
51. Joglekar A, Prjibelski A, Mahfouz A *et al.* A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nat Commun* 2021;**12**:463. https://doi.org/10.1038/s41467-020-20343-5
52. 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. https://doi.org/10.1016/j.celrep.2023.113335

53. Wang C, Shi Z, Huang Q *et al.* Single-cell analysis of isoform switching and transposable element expression during preimplantation embryonic development. *PLoS Biol* 2024;**22**:e3002505. https://doi.org/10.1371/journal.pbio.3002505
54. Mincarelli L, Uzun V, Wright D *et al.* Single-cell gene and isoform expression analysis reveals signatures of ageing in haematopoietic stem and progenitor cells. *Commun Biol* 2023;**6**:558. https://doi.org/10.1038/s42003-023-04936-6
55. Hazzard B, Sá JM, Ellis AC *et al.* Long read single cell RNA sequencing reveals the isoform diversity of plasmodium vivax transcripts. *PLoS Negl Trop Dis* 2022;**16**:e0010991. https://doi.org/10.1371/journal.pntd.0010991
56. 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. https://doi.org/10.1093/qjmed/hcae009
57. 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. https://doi.org/10.1038/s41598-024-72167-8

Briefings in Bioinformatics, 2026, **27**, bbag323
https://doi.org/10.1093/bib/bbag323
Review