Plate-based long-read single cell gene- and isoform transcriptome profiling using scLIS-seq

Koen Deserranno
Ghent University
Elise Callens
Ghent University
Danique Berrevoet
Ghent University
Dieter Deforce
Ghent University
Filip Van Nieuwerburgh
Filip.VanNieuwerburgh@UGent.be

Ghent University

Method Article

Keywords: scRNA-seq, long-read sequencing, Nanopore, isoforms, Smart-seq3xpress, benchmarking

Posted Date: March 21st, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6217988/v1

Additional Declarations: Competing interest reported. K.D. has received travel grants from Oxford Nanopore Technologies (ONT) to present his findings at scientific meetings. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abstract

While contemporary short-read single cell RNA-sequencing allows to decipher tissue composition, discrimination between transcript isoforms remains challenging. Here, we propose single cell long-read isoform sequencing (scLIS-seq), and highlight its performance on Jurkat and HEK293T cells in direct comparison to Smart-seq3xpress (SS3X). scLIS-seq demonstrates sensitive gene and transcript detection with high correlation compared to SS3X and detects at least 10 isoforms of over 2600 genes, while 17.1–21.6% of the reads supported novel isoforms. Direct comparison of the scLIS-seq isoforms to SS3X-reconstructed isoforms demonstrated scLIS-seq's superiority. Overall, scLIS-seq provides a powerful scRNA-seq strategy, enabling long-read transcriptome analysis and isoform detection.

1. Background

Since its conception in 2009, single cell RNA-sequencing (scRNA-seq) has revolutionized the field of transcriptomics, permitting a highly granular view on cellular blood and tissue composition and the downstream construction of human cell atlases (HCA) (1, 2). However, the study of single cell transcriptomes has been focused on the detection and quantification of transcripts for gene expression profiling rather than exploring the variation hidden within and between transcripts. The study of transcript isoforms and transcript heterogeneity within single cells has been hampered by the available methods to explore the transcript's full-length (3). Even though some library preparation strategies such as the 10X Genomics (10X) platform generate full-length cDNA from polyadenylated mRNA transcripts, the cDNA is fragmented in the second stage of the library preparation for compatibility with short-read sequencing (SRS). Resultingly, the obtained sequencing information is highly biased to the 5' or 3' end, and abolishes the opportunity to investigate the presence of internal single nucleotide variants (SNVs) and alternative splicing.

Long-read sequencing (LRS) using Oxford Nanopore Technologies (ONT) or Pacific Biosciences (PacBio) platforms both permit to sequence read lengths over 10 kilobases. However, for long their use in scRNA-sequencing has been hindered by the limited sequencing throughput or low raw read accuracy (4). While an ONT PromethION flow cell could deliver over 100 M reads, a single PacBio SMRT cell on the Sequel II platform would generate 2 M reads. Assuming a minimal sequencing coverage of 20,000 reads per cell for 10X, the latter would only satisfy the sequencing of about 100 cells. Secondly, the raw read quality of the cell barcodes (CB) and unique molecular identifier (UMI) is crucial for accurate cell and transcript deconvolution. During library preparation, each cell is given a unique CB, which allows sequencing of a pool of thousands of cells together. Complimentary, the random UMI sequence compensates for PCR-bias during scRNA-seq library preparation and permits collapsing of all reads with identical CB and UMIs to retrieve quantitative information for downstream differential gene expression profiling. Sequencing errors in the CB and UMI complicate their retrieval and, especially for ONT, were shown to introduce artificial high cell and transcripts counts (5, 6).

As discussed by others, until recently, ONT-sequencing of 10X cDNA required parallel SRS and LRS of the same cDNA library to guide CB and UMI assignment of the LRS results due to the higher error rate in ONT sequencing (7, 8). Alternatively, the scCOLOR-seq strategy adapted the 10X barcode strategy by incorporating predesigned homodimeric nucleotide blocks in the CB and UMI design, to overcome the quality limitations (5, 9). While the former requires additional library preparation and sequencing costs, the latter necessitates the custom design of expensive barcoded beads. Only recently, the raw read accuracy (most relevant for ONT) and throughput (most relevant for PacBio) have increased enough to permit long-read scRNA-seq (LR-scRNA-seq) as a stand-alone strategy, without the need for parallel SRS to guide CB and UMI deconvolution (8). PacBio commercializes the Kinnex single cell protocol, based on first concatenating multiple different 10X cDNA fragments prior to sequencing. The concatenating step is necessary to overcome the lower throughput of their platform compared to SRS, and would result in 80–100M cDNA sequences. For ONT, the recent upgrade to Q20 + chemistry, combined with the latest R10.4.1 PromethION flow cells, and the latest Dorado basecaller would result in 100M reads achieving Q26 simplex raw read accuracy (10–12). At a commonly used minimum sequencing depth of 20,000 reads per cell for droplet-based single cell approaches, this would theoretically result in the possibility to sequence 5,000 cells per flow cell.

At this point, the number of scRNA-seq kits compatible with LRS remains limited. Both PacBio and ONT start from cDNA prepared by the commercial 10X platforms or the split-pool ligation-based transcriptome sequencing (SPLiT-seq) technology of Parse Biosciences (13, 14). While 10X permits high-throughput profiling of thousands of cells, dedicated microfluidic devices are required for droplet generation. Both 10X and Parse Biosciences are designed to start from substantial input cell numbers, while in clinical settings only a small or rare cell population can be of interest. Therefore, this results in a high cost per cell if prepped using either of these platforms. Moreover, neither strategy allows for direct pairing of the phenotype information obtained by indexed Fluorescence Activated Cell Sorting (FACS) directly to the single cell transcriptome data. Finally, both methods show lower sensitivity compared to miniaturized plate-based scRNA-seq strategies, e.g. FLASH-seq and Smart-seq3(xpress) (SS3(X)) (15, 16). Interestingly, the plate-based SS3(X) strategies employ a clever approach based on *in silico* stitching of SRS reads with the same unique UMI to reconstruct the original transcript based on alternative 3’ ends of the tagmented reads. However, it’s accuracy has been debated (17).

To overcome these limitations, we developed single cell long-read isoform sequencing (**scLIS-seq**). Starting from full-length cDNA as prepared using the well-established plate-based SS3X protocol, we developed a fast novel strategy for single cell barcoding and direct sequencing of full-length cDNA amplicons on the ONT PromethION platform. We performed scLIS-seq on Jurkat and HEK293T cells to demonstrate our strategy’s performance and validated it against the SRS-SS3X data derived from the identical cDNA. We adapted and benchmarked the ONT wf-single-cell and scywalker bioinformatics workflows for compatibility with our scLIS-seq constructs, permitting downstream gene expression profiling without the need for parallel SRS. Subsequently, we focused on scLIS-seq’s isoform identification capabilities and compare the identified isoforms with these reconstructed by UMI-linking in SS3X.

2. Results and Discussion

2.1. scLIS-seq library construction

Two major adjustments have been made compared to the regular SS3X protocol. First, in contrast to the 10X and Parse Biosciences workflow in which cDNA barcoding is performed at the start of the library preparation, barcoding in the regular SS3X protocol is only performed in the last PCR-step to retain full-transcript coverage. In SS3X, dual unique index primers hybridize to the 5' Tn5-motif in the template-switching oligo (TSO) and anchors introduced by the Tn5 transposase. However, to retain full-length cDNA amplicons in scLIS-seq, the transposase step is omitted, cancelling the second primer hybridization site. To enable the barcoding of the cDNA molecules in each well, we moved from dual index barcoding to single index barcoding, as only the i5 barcode can hybridize to the untagmented SS3X cDNA near the TSO. As a reverse primer, we designed a new oligonucleotide (scLIS-seq-RP) complementary to the oligodT-anchor sequence at the 3' end of the transcript to preserve the full-length capabilities. As the unique dual index primers often come premixed, we optimized the PCR-reaction and melting temperature of the reverse primer to outcompete the i7 barcode if present (Supplementary Materials, Fig. 1). Adapting the SS3X protocol in this way, while keeping the cDNA unaltered, allows for parallel sequencing of the same library on both SRS as LRS platforms. The scLIS-seq library construction strategy is illustrated in Fig. 1(18).

Secondly, to compensate for the lower quality at the beginning and end of each ONT read, spacer sequences were introduced. Upon attachment and detachment of the ONT motor protein to the nanopore, the read quality drops. While the overall read quality is advertised to be over Q20, quality scores at both ends of the read are typically lower (Supplementary Materials, Fig. 2). This impacts the identification of full-length reads. Full-length reads are defined as reads having the 5' TSO adapter and 3' oligo-dT sequences. The lower quality ends of the read interfere with the identification of these adapters. Therefore, we added an additional 22 nucleotide long spacer sequence to scLIS-seq-RP, similarly to what ONT implemented for their 10X protocol. However, at the 5' TSO side, we could not add such a spacer next to the i5 barcode sequence without losing direct SRS compatibility. Therefore, ONT native barcode ligation was performed. By introducing ONT barcodes at either end of the scLIS-seq library molecules, even when only a single sample per flow cell is sequenced, additional buffering nucleotides are introduced to compensate for the lower quality at the start of the read (19).

2.2. Gene expression profiling

To evaluate the performance of scLIS-seq for single cell transcriptomics profiling, a direct comparison between SR-scRNA-seq and LR-scRNA-seq was performed. Starting from the identical cDNA of 96 individual Jurkat and 144 individual HEK293T cells, libraries for SS3X and scLIS-seq were prepared in parallel and sequenced (Supplementary Materials, Table 1). scLIS-seq yielded a total of 30.8 M and 52.8 M raw long-reads for the Jurkat and HEK293T cells, respectively, compared to 28.5 M and 190.4 M short-reads for SS3X. For scLIS-seq, after quality and adapter filtering, 15.3 M and 19.4 M long-reads were retained. The median read length was 926 nt and 687 nt for Jurkat and HEK293T, respectively. In

contrast to 10X-based data-analysis strategies, the exact CBs used during library preparation are known upfront in both scLIS-seq and SS3X and do not have to be inferred from the sequencing data based on a provided whitelist containing millions of potential CBs. CB calling from such extensive list of potential CBs has been shown to introduce uncertainty in CB identification (20, 21).

scLIS-seq data analysis was performed by running both the wf-single-cell and scywalker data-analysis pipelines. At the moment of writing, these strategies differ in several key elements. First, wf-single-cell uses a read splitting step to split fused reads (22). Concatenated fused reads can arise during nanopore library preparation because of the ligation of cDNA molecules to another. Based on the presence of adapter sequences, wf-single-cell splits a long, fused reads into individual subreads. Scywalker has yet to implement such step. Secondly, wf-single-cell uses a correction algorithm to account for sequencing errors in the UMI. Based on initial clustering of the reads by assigned gene or genomic intervals, UMIs up to two Levenshtein distances apart are corrected into one.

scLIS-seq demonstrated sensitive gene and UMI detection, on par with the results obtained with regular SR-SS3X (Fig. 2, A, B, D, E), using both the wf-single-cell and scywalker pipelines. The scywalker pipeline detected slightly more unique molecules per cell than wf-single-cell. This might be due to the lack of a UMI-correction step in scywalker as discussed above. Strikingly, these results were obtained with significantly lower reads per cell than SS3X (Fig. 2, C & F). The mean number of raw reads per cell in SS3X is at least three times higher than for scLIS-seq. However, in SS3X, the existence of both 5' UMI-containing reads as internal non-UMI-reads requires deeper sequencing to capture the complete library complexity, as illustrated before (15). Thanks to the long-reads and the omission of the tagmentation step in scLIS-seq, all library fragments contain the UMI-sequence, allowing sequencing at lower depth with similar performance. To note, despite having been provided with the same number of raw input reads, wf-single-cell retrieves less reads per cell compared to scywalker (Supplementary Materials, Fig. 3). This might be attributed to differences in the stringency and filter criteria between both analysis pipelines.

scLIS-seq demonstrated high correlation in the number of genes and UMIs detected per cell (Fig. 2, G & H) in comparison to SS3X. Both for Jurkat and HEKT293T, we found significantly high correlation in the number of UMIs detected per cell between scLIS-seq and SS3X (Jurkat: Pearson's r = 0.83; HEK293T: Pearson's r = 0.84, based on wf-single-cell analysis), as illustrated in Fig. 2, I & J. These correlation coefficients are slightly lower than these observed by scNanoGPs for the A-375 and H203 cancer cell lines (r = 0.97 and r = 0.89, respectively), albeit possibly impacted by the number of cells profiled (20). Furthermore, we compared the normalized single cell gene expression results between SR SS3X and scLIS-seq for the commonly identified genes. We detected high gene expression correlation coefficients for Jurkat and HEK293T (Fig. 2, K & L). The high concordance between both strategies resulted in a Pearson correlation coefficient of 0.88 and 0.91 for Jurkat and HEK293T, respectively, comparable with the results from scNanoGPS (20). This demonstrates the capability of scLIS-seq to serve as a stand-alone technology for single cell transcriptome profiling.

Finally, to investigate if the UMI-counts for scLIS-seq are not artificially inflated due to ONT-sequencing errors in the UMI-sequence, we compared the extent of UMI-overlap between SR SS3X and scLIS-seq (Fig. 2, M & N). Crucially, as UMIs are not unique within a single cell (UMI-length of 8 nucleotides results in 65,536 theoretical combinations), we additionally accounted for the chromosome the read aligned to. We therefore compared the overlap of the chromosome – cell barcode – UMI (chr-CB-UMI) combinations between SS3X and scLIS-seq in the Jurkat cells. We found that 48.92% and 69.65% of these combinations are unique to scLIS-seq, respectively for the wf-single-cell and the scywalker pipeline. As discussed before, the lower level of overlap in scywalker compared to wf-single-cell might be due to the lack of a UMI-correction step. In a recent preprint a similar comparison was performed for 10X generated cDNA sequenced on Illumina and PacBio, and retrieved that 10–30% of combinations were unique to PacBio (23). However, they did not take the chromosome information into account and are therefore potentially biased to higher overlap. To investigate this further, we compared the cell with the highest number of reads in the scLIS-seq data of the Jurkat cells (262,336 reads) to its SS3X counterpart (629,586 total reads). Upon profiling the level of overlap for these cells, we found that the number of chr-CB-UMI combinations unique to scLIS-seq decreased to 38.82% (Supplementary Materials, Fig. 4A). We also evaluated if further filtering by only selecting those UMI-sequences in which each base has at least Q30 accuracy increased the extent of overlap. The number of scLIS-seq unique combinations upon applying this additional filter only slightly improved to 35.14% (Supplementary Materials, Fig. 4B). Therefore, we conclude that sequencing depth and not UMI-accuracy is the most important limitation for using LRS for single cell transcriptomics. This corroborates the findings of previous research that ONT as stand-alone technology delivers sufficiently accurate information for single cell transcriptomics, however sequencing throughput remains the major bottleneck (21, 22).

2.3 Isoform identification using scLIS-seq

We further evaluated to which extent scLIS-seq could resolve distinct transcript isoforms. Considering the reads with CBs that are also present in the filtered count matrix, isoforms outputted by Isoquant as implemented in scywalker were characterized. 50.61% and 36.39% of the reads obtained from the Jurkat and HEK293T cells, could be assigned to single unique isoform (Supplementary Materials, Fig. 5), respectively. For more than 2633 and 2652 genes, at least 10 unique isoforms per gene were detected for Jurkat and HEK293T, respectively (Fig. 3, A). We further classified reads corresponding to a unique isoform in their corresponding SQANTI3-like isoform class. (Fig. 3, B). For Jurkat cells, 32.82% of the these reads could be classified as a full splice match (FSM), reflecting reads having the corresponding number of exons and splice junctions compared to the Ensembl reference (24). This percentage is slightly lower compared to the 37–39% recently observed for ONT cDNA and PacBio MAS-ISO-seq (the precursor of PacBio Kinnex) using 10X Genomics single cell cDNA from lung tissue (25). In contrast, only 12.23% of the reads was classified as FSM for HEK293T, while 32.44% was classified as incomplete splice match (ISM). ISMs are characterized by reads having fewer exons than the reference, with missing exons at the outer side of the transcript, but with matching splice junctions compared to the reference (24, 26). Both results corroborate the findings of previous research, demonstrating that most reads do not represent complete isoforms due to library preparation artifacts and RNA degradation, as further

discussed below (26, 27). To note, 17.1 and 21.6% of the reads supported isoforms annotated as novel-in-catalog (NIC) or novel-not-in-catalog (NNC), for HEK293T and Jurkat, respectively.

For each of the unique isoforms annotated as NIC and NNC, we assessed the number of cells supporting that isoform (Fig. 3, C and D). While theoretically it could be expected that most of the clonal Jurkat and HEK293T cells support each isoform, dropouts due to variations in sequencing depth are to be expected, especially for lowly expressed isoforms (28, 29). Alternatively, novel isoforms expressed by only a small percentage of these clonal cells are probably artifacts, as further discussed below. In addition, we evaluated the impact of cell cycle stage as a potential confounder. Therefore, we profiled the number of unique novel isoforms according to cell cycle phase (Methods). The identified cell cycle specific isoform expression for the novel isoforms is illustrated in Fig. 3, E and F. Specifically, our results enable to study cell cycle dependent isoform expression of key mitotic genes, e.g. *AURKB.* In literature, the promotion of an intron-retention isoform during the S/G2 phase was suggested (30, 31). Specifically, we calculated the number of cells expressing *AURKB* isoforms annotated as novel in each cell cycle phase, and found their almost exclusive expression of these isoforms in S and G2M phase. For the novel isoforms with ENST00000584561 or ENST00000583124 (Ensembl biotype: Intron retention) as closest known isoform, we show their exclusive expression in the cells within the S and G2M-phase (Supplementary Materials, Fig. 6). Overall, these results show that scLIS-seq enables to identify potential novel isoforms and their relative expression within each cell cycle stage.

RNA-sequencing library preparation is sensitive to several potential artifacts (32). One of the main artifacts, which could interfere with isoform detection, appears during reverse transcription (RT) due to internal mRNA priming or template switching (TS). Internal mRNA priming can occur due to oligodT-primer hybridization with poly(A)-stretches within the transcript (33, 34). In TS artifacts, the newly synthesized first strand of cDNA can hybridize with another RNA molecule during reverse transcription, and use the former strand as a template (35). These RT-artifacts are not unique for scLIS-seq, but might be hidden in the SRS data. Additionally, the template-switching-oligo (TSO) used during library preparation can prime internally on the mRNA or first-strand cDNA molecules (25, 32). Furthermore, the emergence of novel transcription start sites (TSS) and transcription termination sites (TTS) are hard to distinguish from RNA-degradation (26).

To study these artifacts, we passed the scywalker generated transcriptome file of the HEK293T cells to SQANTI3 (26). As SQANTI3 does not keep single cell level information, conclusions are based on the total number of isoforms identified in the Jurkat and HEK293T cells. Based on the SQANTI3-output, we profiled the number of presumable RT-switching events for each isoform classification category in the HEK293T-cells. Figure 4A illustrates the fraction of isoforms with underlying RT-switching potential in each classification category for the HEK293T-cells. Strikingly, even for the FSM annotated isoforms, SQANTI3 reported that 12.8% could be potential RT-artifacts. To note, older versions of SQANTI3 (until version 3.5.2, released January 2025) contained a bug in the RT-switching counting step, which underestimated the number of RT-artifacts, and complicates comparison of these results with previous research (SQANTI3 v3.5.0 resulted in 3.7% RT-artifacts for FSM isoforms). Recent analysis of the TCGA

database illustrated that up to 75% of the reported exitrons, i.e. introns with protein coding exons, show hallmarks of artificial RT-switching, highlighting that also currently reported refence isoforms are vulnerable to RT-bias (35). Only direct RNA-sequencing allows for unbiased discrimination between false and real transcripts with RT-switching potential. Complimentary, to evaluate the potential of oligodT-primer mediated internal mRNA-priming, the percentage of adenosine nucleotides in the 20 nt window downstream of the TTS was evaluated in Fig. 4B. SQANTI3 recommends to filter out transcripts with over 60% adenosine residues, which would correspond to 3.03% of all isoforms identified in the HEK293T-data.

Finally, we checked the presence of the TSO-mediated strand invasion events in the identified isoforms (15, 16). Therefore, we checked the presence of the 'WWGGG' TSO-motif in the 20-nucleotide genomic window upstream of the TSS of each isoform identified by SQANTI3 (Fig. 4, C). Overall, we found that in 5.59% of the classified isoforms, a potential strand invasion event cannot be excluded. We also repeated this analysis on a subset of 1 M long reads obtained from A-375 cells prepared using 10X Genomics 3' GEX chemistry as performed by the authors of scNanoGPS (20). Across all their isoforms, 0.97% possessed the 'ATGGG' TSO-motif in their 20-bp upstream region (Fig. 4, D). Importantly, the main artifacts in 10X Genomics 3' chemistry, annotated as TSO-TSO artifacts, were already removed during sample preparation in scNanoGPS using biotin-streptavidin pull-down (5, 20). The TSO-TSO artifacts are characterized by the presence of the TSO sequence followed by its reverse compliment in the same read. If left untouched, these artifacts have shown to make up 30–50% of all reads sequenced from the 10X Genomics libraries (5). Together with the intermediate purification step between RT and cDNA pre-amplification during 10X library prep to remove excess TSO, this artifact removal step might explain the lower fraction of TSO-mediated strand invasion generated isoform events observed.

While in scLIS-seq such artifact removal step was not performed, it could be implemented if beneficial. We tested the presence of TSO-TSO artifacts in a subset of the raw HEK293T scLIS-seq sequencing data without artifact removal. Based on this analysis, 9.05% of the reads could potentially be TSO-TSO artifacts. However, this number might be an overestimation for scLIS-seq due to the presence of fused long-reads. As discussed before, fused long-reads can arise due to unintended ligation of different cDNA to each other during the ligation steps in the ONT protocol. It is therefore possible that two consecutive TSOs in one read originate from 2 different cDNAs rather than a TSO-TSO artifact. Fused long-reads are not unique to scLIS-seq. This is exemplified by the finding that 7.99% of the raw reads from the A-375 subset, despite the artifact removal, still showed the presence of two or more consecutive TSOs per read. During processing by wf-single-cell, these fused reads are split into their individual segments before further processing (22). However, this step has not been implemented in scywalker yet. In summary, while adequate curation of the isoforms identified using scLIS-seq is recommended, as is also the case for other strategies, scLIS-seq facilitates the detection of both known and novel isoforms.

2.4 Isoform Reconstruction: UMI-Stitching in SS3X vs. scLIS-seq

Both Smart-seq3, SS3X, and FLASH-seq offer the unique feature to partially reconstruct the full-length of individual transcripts using short-reads based on UMI-linkage. Upon successful cDNA pre-amplification, each cDNA copy will be randomly fragmented and tagged in the tagmentation process, generating both 5’ UMI-reads as internal reads. Additionally, some cDNA copies, all originating from a single mRNA-transcript having a single UMI, will now have different 3’ ends, while their 5’ end remains identical. If these reads are paired-end sequenced, the different 3’ ends can be stitched together if their UMI at the 5’ end is the same (15, 36). However, others noted ambiguity in these reconstructions and found up to 43% error rate in assigning reconstructed transcripts to specific known SIRV isoforms (17).

Here, we provide a direct comparison of the identical cDNA molecules sequenced using scLIS-seq and these sequenced and reconstructed using SR SS3X of the Jurkat cells. For comparison, we extracted reads with identical CB-UMI combined sequences, aligned to the same gene, from both strategies. As SS3X and scLIS-seq start from the identical cDNA, strand invasion artifacts caused by TSOs infiltrating within the transcript body rather than on the 5’ end, are expected in both strategies. For SS3X, these artifacts were priorly filtered out from the aligned .bam file (16). Of note, stitching was performed using the uncorrected UMI-tag (UX) instead of the corrected UMI (UB) as outputted by zUMIs, as coincidental false stitching of two UBs but with different UX was observed (Supplementary Materials, Fig. 7).

UMI-stitching of the Jurkat SS3X data resulted in 1,519,534 transcript reconstructions longer than 300 nucleotides across all cells. From these, we searched for their scLIS-seq counterparts with identical CB, UMI, and chromosome alignment. After filtering to remove UMI sequences consisting of homopolymer adenosine residues and selecting these pairs with concordant gene assignments between the reconstructed transcripts and scLIS-seq, 563,961 pairs were obtained. We profiled the common pairs from each strategy using SQANTI3. Detailed examples of the short-reads, reconstructed reads, and scLIS-seq reads of such a pair can be found in Supplementary Materials, Fig. 7.

As expected, for 96.75% of these pairs, the scLIS-seq corresponding read is longer than the reconstructed transcript. scLIS-seq demonstrates a broader isoform length distribution than the isoforms obtained using UMI-stitching (Fig. 5, A). Classification of the identified isoforms in their respective categories revealed that most isoforms reconstructed by stitcher were annotated as NNC (Fig. 5, B), which can be attributed to the partial nature of their reconstruction. Depending on the number of tagmented positions within a single cDNA molecule, individual exons can only be partially sequenced or even completely missed, and hence are absent from the reconstructed transcript. Consequently, SQANTI3 classifies these false boundaries as novel splice donor/acceptor sites. Indeed, upon profiling the splice site junctions in both strategies, we found that 42.2% of the NNC isoforms reconstructed by stitcher have non-canonical splice junctions, validating our hypothesis (Fig. 5, C).

Apart from the mainly partial nature of the SS3X-reconstructions, further in-depth analysis revealed several additional shortcomings of reconstructing isoforms based on UMI-stitching (Supplementary Materials, Fig. 8). Both remaining strand invasion artifacts as well as false stitching of identical UMIs with spurious 5’ starting sites within the same CB, cause interference in the SS3X-reconstructions

compared to the long-read counterpart. The latter might be attributable due to UMI-collision. Overall, we conclude that SS3X-derived isoform-reconstructions should be handled with caution, and are, both in transcript length as in accuracy, inferior to direct profiling using scLIS-seq.

3.5 Limitations

Within the field of long-read single cell transcriptomics, a clear need for alternative library preparation strategies and benchmarking of these strategies with their respective analysis pipelines, was expressed (8). scLIS-seq offers as a complimentary strategy for LR-scRNA-seq compared to the existing 10X strategies. It offers the sensitivity of the miniaturized SS3X protocol with long-read capabilities for isoform. The compatibility with indexed FACS-sorting allows to correlate the phenotype observed directly with its transcriptome and isoform landscape highlights one of its unique advantages. The most limitations of the current version of scLIS-seq are common to all oligo(dT) and TSO-based scRNA-seq strategies. The fact that most reads are actually not full length, and the presence of miscellaneous artifacts warrant further RT-optimization and the development of novel strategies. Specifically for scLIS-seq, optimization is possible in improving the sequencing yield of true cDNA molecules obtained. A persistent library preparation artifact, which arises from direct hybridization of the TSO with the oligo(dT)-primer without cDNA insert, limits throughput. In the current implementation of scLIS-seq, the artifact remained abundant in the sequencing data despite attempts to remove it using agarose gel electrophoresis and subsequent clean-up.

Albeit a limited number of cells was used to showcase the performance of scLIS-seq, and further extensive benchmarking of these methods remains warranted, this small cell population mimics the situation in the clinic. In the case of oncology, a limited number of cells can be isolated based on a combination of specific cell surface markers. scLIS-seq allows isoform profiling of this remaining small cell population in a cost-effective plate-based manner, upon FACS-isolation in individual wells.

3. Conclusion

In conclusion, we have demonstrated that scLIS-seq provides a powerful plate-based strategy for single cell transcriptome profiling using long-read sequencing. Our results reveal a high correlation in gene expression between scLIS-seq and short-read SS3X while highlighting the importance of sequencing throughput for comparable UMI-detection. Furthermore, we demonstrate scLIS-seq’s superior performance to identify transcript isoforms, surpassing that of short-read SS3X reconstructions. We explicitly show that scLIS-seq allows the retrieval of longer and more accurate isoforms compared to short-read reconstructions using UMI-stitching. Overall, scLIS-seq offers a LR-scRNA-seq, with the advantages of direct compatibility with FACS-based single cell isolation, reliable cell barcode identification, and sensitive isoform detection.

4. Materials and methods

4.1 Cell culture and isolation

Jurkat T lymphoblasts (clone E6-1) were purchased from the DSMZ repository (Braunschweig Germany) and cultured in RPMI 1640 medium (Gibco) supplemented with 10% fetal bovine serum (FBS, Hyclone) and 2 mM L-glutamine (Gibco) at 37°C and 5% CO_2. The HEK293T kidney cells were cultured in DMEM (Gibco, Thermo Fisher Scientific, Waltham, MA, USA). Jurkat T lymphoblasts were first blocked with anti-human FcR (Miltenyi, #130-059-901) for 10 minutes at room temperature to avoid non-specific binding of antibodies. Next, these cells were incubated with antibodies against CD3 (clone UCHT1, BioLegend cat. Nr. 300440, FITC, dilution: 1:200) and TCR α/β (clone IP26, BioLegend cat. Nr. 306718, APC, dilution: 1:200). Both Jurkat and HEK293T cells were washed using phosphate buffered saline (PBS, Gibco), and finally were resuspended in PBS. Propidium iodide staining was performed to mark dead cells. Additionally, for Jurkat, gating for living CD3 + TCRαβ + cells was performed. The cells were sorted as single cells into 384-well Armadillo plates plate (Thermo Fisher Scientific, Waltham, MA, USA) containing 0.3 µL SS3X lysis buffer and 3 µL of 5 cSt silicon oil overlay (Merck) per well. Sorting was performed using BD FACSAria Fusion (BD Biosciences) for Jurkat and BD FACSAria II (BD Biosciences) for HEK293T with a 70-µm nozzle. After sorting, each plate was immediately sealed, centrifuged, and stored at -80°C upon further processing.

4.2 Preparation of single cell cDNA

After FACS, cDNA was generated on single-well level using the well-established SS3X protocol, according to the V2 iteration as published on protocols.io (37). All reagent nanodispensing steps were performed using the I.DOT liquid handler (Dispendix). The pre-amplification PCR was performed for 12 cycles. To each well, 9 µL of nuclease-free H_2O (VWR, Radnor, PA, USA) was added to create 1:10 cDNA dilution.

4.3 Regular SS3X library preparation for comparison with scLIS-seq

Modifications compared to the V2 published version of the protocols.io protocol include adjusting the input of Tn5 transposase enzyme (Diagenode, Luik, Belgium) to 0.014 µL per reaction for Jurkat and 0.0038 µL TDE1 (Illumina, San Diego, CA, USA) for HEK293T. Additionally, the use of 0.2% SDS to stop the tagmentation reaction and the inclusion of 0.025% Tween-20 within the index PCR mix to counteract the effects of SDS were omitted. Furthermore, the index PCR mix was refined by utilizing only 0.1 µM of each index primer (IDT for Illumina UD Indexes, Integrated DNA Technologies, Coralville, IA, USA), and 0.750 ng of tRNA carrier (Thermo Fisher Scientific, Waltham, MA, USA) was added to each reaction to enhance library yield. The complete list of index primers used can be found in Supplementary Files, Tables 2 & 3. The final PCR-reaction was performed for 15 cycles for Jurkat and 14 cycles for HEK293T. Purification was performed using AMPure XP Beads (Beckman Coulter, Brea, CA, USA) in a 0.7:1 beads-to-sample ratio. The SS3X libraries were sequenced on AVITI™, generating PE150 reads.

4.4 scLIS-seq library preparation

After centrifugation, 1 µL of the diluted cDNA was transferred to a new 384-well Armadillo plate (Thermo Fisher Scientific, Waltham, MA, USA) and stored on ice. To each well, 2.75 µL of PCR-mix containing 1 X of 5 X Phusion PLUS Buffer (Thermo Fisher Scientific, Waltham, MA, USA), 0.20 mM dNTPs/each

(Thermo Fisher Scientific, Waltham, MA, USA), 0.25 µM reverse primer (/5Phos/TTT CTG TTG GTG CTG ATA TTG CGC ATC AGC AGC ATA CGA, HPLC purified, Integrated DNA Technologies, Coralville, IA, USA), 0.05 µL Phusion PLUS DNA polymerase, was added, and adjusted to 2.75 µL with nuclease-free H_2O. Subsequently, on a per-well level, 0.125 µM of pre-mixed unique dual index primers (IDT for Illumina UD Indexes, Integrated DNA Technologies, Coralville, IA, USA) was added to achieve a total reaction volume of 5 µL. The complete list of index primers used can be found in Supplementary Files, Tables 2 & 3. All volume transferring steps to the 384-well plate were performed using the FAST Liquid Handler (Formulatrix, Dubai, UAE). Thermal cycling was performed starting with initial denaturation for 30 seconds at 98°C, followed by 24 cycles of 10 seconds 98°C, 30 seconds 64°C, and 2 minutes at 72°C. Final elongation was performed for 5 minutes at 72°C. After completion of the PCR, all wells were pooled and purified using AMPure XP beads (Beckman Coulter, Brea, CA, USA) in 1:1 bead-to-sample ratio. The purified library was run on a 2% Agarose E-gel (Thermo Fisher Scientific, Waltham, MA, USA). All fragment sizes above 300 nucleotides were excised and extracted from the gel using Zymoclean Gel DNA Recovery kit (Zymo research, Irvine, CA, USA) according to the manufacturer's protocol. The resulting purified library was quantified using the Qubit dsDNA assay (Thermo Fisher Scientific, Waltham, MA, USA).

4.5 ONT library preparation

ONT library preparation was performed using the Native Barcoding Kit SQK-NBD114.24 (ONT, Oxford, United Kingdom). Briefly, 110.4 ng (Jurkat) and 139.5 ng (HEK293T) sequencing library from the previous step was mixed with 1.75 µL Ultra II End-prep Reaction Buffer (NEB, Ipswich, MA, USA), 0.75 µL Ultra II End-prep Enzyme mix (NEB, Ipswich, MA, USA), and 1 µL of nuclease-free H_2O, prior to AMPure XP bead purification (Beckman Coulter, Brea, CA, USA) in a 0.7:1 bead-to-sample ratio. Native barcode ligation was performed using 7.5 µL of end-prepped DNA, 2.5 µL of native barcode (ONT, Oxford, United Kingdom), and 10 µL of Blunt/TA Ligase Master Mix (NEB, Ipswich, MA, USA). After 20 minutes of incubation at room temperature, 2 µL of EDTA (ONT, Oxford, United Kingdom) was added and the resulting volume was again purified in a 0.7:1 bead-to-sample ratio. Lastly, 30 µL of the barcoded sample was mixed with 5 µL of Native Adapter (ONT, Oxford, United Kingdom), 10 µL of 5X NEBNext Quick Ligation Reaction Buffer (NEB, Ipswich, MA, USA) and 5 µL of Quick T4 DNA ligase (NEB, Ipswich, MA, USA). After 20 minutes of incubation, Short Fragment Buffer (ONT, Oxford, United Kingdom) was used for purification along with a 0.7:1 bead-to-sample ratio. After purification, 30 µL of the final sequencing library (Jurkat: 118 fmol, HEK293T: 80 fmol) was mixed with 68 µL of Library Beads (ONT, Oxford, United Kingdom), and 100 µL of Sequencing Buffer (ONT, Oxford, United Kingdom) prior to loading on the R10.4.1 PromethION flow cell.

4.6 ONT basecalling and quality filtering

Basecalling was performed with Dorado (v.0.8.3) using the superhigh accuracy model. Chopper (v.0.7.0) (38) was used for quality filtering to retain reads having a Q-score above 10. Subsequently, primer dimers, adapter dimers, and the forward primer – TSO – oligodT – reverse primer artifact was removed using Chopper with flag -l 350.

4.7 SS3X data-analysis

The raw AVITI .fastq files were processed using zUMIs v.2.9.7e (39). The Ensembl GRCh38 release 111 was used as a reference. As barcode_file, the combinations of the i7 and i5 indices stated in Supplementary Files, Tables 2 & 3 was used.

4.8 Adaptation of wf-single-cell for scLIS-seq

The ONT wf-single-cell nextflow pipeline was used for processing of the quality filtered scLIS-seq data (40). However, this pipeline is built for processing 10X cDNA, and therefore expects that the CB and UMI are positioned next to each other. For scLIS-seq, we made several adaptations to the multiple scripts of this pipeline to adjust the CB and UMI extraction procedure and provided a custom barcode whitelist. The adjusted pipeline can be found in the Supplementary Materials - Code. A custom 10X-like reference genome was generated based on GRCh38 release 111 using Cell Ranger mkref (41).

4.9 Adaptation of scywalker for scLIS-seq

Equivalently, also scywalker (v0.110.0) is built to be compatible with 10X (27). The scywalker scripts were minimally adjusted by changing the adapter sequence to 'CGGCGACCACCGAGATCTACAC' and the set_umi command (Supplementary Materials, Code). Scywalker was run with flags -v 2 -d 20 -refdir {refdir_hsa111} -sc_whitelist {whitelist} -sc_expectedcells 70 (jurkat) or 144 (HEK293T) -sc_umisize 8 -sc_barcodesize 10 -sc_adaptorseq CGGCGACCACCGAGATCTACAC -samplesheet {samplesheet}. The scywalker reference genome was generated running scywalker_makerefdir -organelles chrMT, providing Ensembl GRCh38 release 111 as input.

4.10 Downstream analysis in Seurat

The output of zUMIs, wf-single-cell, and scywalker were processed by Seurat 5.0.3 (42–46). Specifically, for zUMIs, the Smartseq3xpress.dgecounts.rds was used. For wf-single-cell and scywalker, gene_raw_feature_bc_matrix and sc_gene_counts_raw-isoquant_sc-sminimap2_splice-Jurkat_SS3X_plateC.10x were used. The Seurat count objects were filtered to only retain cells with at least 300 features, expressed in at least 3 cells, with at least 20,000 reads but not more than 1,000,000 reads. biomaRt (2.58.2) was used for the interconversion between gene symbols and gene IDs (47, 48). Cell cycle phase was assigned using Seurat's CellCycleScoring, based on built-in cell cycle markers (49). Gene expression was normalized using LogNormalize with scale.factor = 10,000. For construction of the venn diagrams with overlapping chromosome_CB_UMI between strategies, Rsamtools (2.18.0) (50) and GenomicAligments (1.38.2) (51) were used to extract the chromosome, CB, and UMI tags, from the respective .bam files. The venn diagram was constructed using the VennDiagram (1.7.3) package (52).The ggplot (3.5.1) and ggpubr (0.6.0) packages were used for constructing figures (53, 54). For illustration of the genome tracks, the Gviz package (1.46.1) was used (55).

4.11 Identification of isoform artifacts using SQANTI3

For the isoform classification, the read_assignments-isoquant_sc-sminimap2_sample_.tsv files, as outputted by scywalker, were used and loaded into R. For construction of the upset plots, the UpSetR package (1.4.0) was used (56). For the profiling of potential artifacts in the isoforms, the isoforms-isoquant_sc-sminimap2_splice-HEK293T_SS3X_144.gtf was used as input for SQANTI3 (5.3.6) (26). The Homo Sapiens Ensembl release 111 was used as reference. SQANTI3 was run with the following parameters: python ~/tools/sqanti3_qc.py {scywalker_HEK293T_gtf} {scywalker_reference_gtf} {scywalker_reference_fasta} --polyA_motif_list {SQANTI3_public_polyAmotif} -o HEK293T_scywalker_sqanti3 -d {SQANTI_analysis_dir} -t 20 --report both –isoAnnotLite. For profiling of the TSO-mediated strand invasion within the isoforms, the SQANTI3 code was adapted (Supplementary Materials, Code). Plots were made based on the SQANTI3_report.R code as provided in their GitHub. For identification of TSO-TSO artifacts, Pychopper (v2.7.10) (57) was used to assess the presence of the TSO and oligo(dT)-primer sequences in each read. A read was classified as having a TSO-TSO artifact if the read consisted of a TSO followed by its reverse complement.

4.12 SS3X isoform reconstruction and comparison with scLIS-seq

Transcript reconstruction based on the SS3X Jurkat data was performed using stitcher.py. (58) As input .bam file, the zUMIs generated UBcorrected.sorted.bam was first filtered to remove strand invasion artifacts using the filterInvasionEventsfromBAM.R code from the FLASH-seq paper, adapted to account for the WW-spacer and 3’ GGG motif (16). The wf-single-cell generated .bam file was deduplicated using umitools dedup (59). Using custom python scripts, the intersecting CB-UMI-gene pairs were extracted from the scLIS-seq and stitcher reconstructions. The .fastq files corresponding to these pairs were extracted from the respective .bam files using samtools bam2fq, and used as input for SQANTI3 (v.5.3.6) specifying the .fastq input format using --fasta.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and analysed during the current study are available in the ENA repository, using accession number PRJEB86878.

Competing interests

K.D. has received travel grants from Oxford Nanopore Technologies (ONT) to present his findings at scientific meetings. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

K.D. and this research are supported by the Special Research Fund (Bijzonder Onderzoeksfonds, BOF, University Ghent, BOF21/DOC/042) website: https://www.ugent.be/en/research/funding/bof. The authors declare no further external financial support was received for the research, authorship, and/or publication of this article.

Authors' contributions

K.D. Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing-original draft; **E.C.** Investigation, Methodology, Writing – review & editing; **D.B.** Investigation, Methodology, Writing – review & editing; **D.D.** Funding acquisition, Writing – review & editing; **F.V.N.** Conceptualization, Methodology, Supervision,Funding acquisition, Writing – review & editing.

Acknowledgements

We thank Sarah De Keulenaer, Ellen De Meester, and Sylvie Decraene of the Ghent University NXTGNT sequencing core facility for their expertise and help in performing the Element AVITI sequencing runs. We thank Thomas Michiels, Laurenz De Cock, Tamara De Vos, and the Flow Cytometry Core Facility at Ghent University for support with sample sorting.

References

1. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods. 2009;6(5):377-82.
2. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The human cell atlas. elife. 2017;6:e27041.
3. Pan L, Dinh HQ, Pawitan Y, Vu TN. Isoform-level quantification for single-cell RNA sequencing. Bioinformatics. 2022;38(5):1287-94.
4. Shiau C-K, Lu L, Kieser R, Fukumura K, Pan T, Lin H-Y, et al. High throughput single cell long-read sequencing analyses of same-cell genotypes and phenotypes in human tumors. Nature communications. 2023;14(1):4124.
5. Lebrigand K, Magnone V, Barbry P, Waldmann R. High throughput error corrected Nanopore single cell transcriptome sequencing. Nature Communications. 2020;11(1):4025.
6. Mikheenko A, Prjibelski AD, Joglekar A, Tilgner HU. Sequencing of individual barcoded cDNAs using Pacific Biosciences and Oxford Nanopore Technologies reveals platform-specific error patterns.

Genome Research. 2022;32(4):726-37.
7. Xu Z, Qu H-Q, Chan J, Kao C, Hakonarson H, Wang K. Single-Cell Omics for Transcriptome CHaracterization (SCOTCH): isoform-level characterization of gene expression through long-read single-cell RNA sequencing. bioRxiv. 2024:2024.04.29.590597.
8. Gupta P, O'Neill H, Wolvetang EJ, Chatterjee A, Gupta I. Advances in single-cell long-read sequencing technologies. NAR Genomics and Bioinformatics. 2024;6(2):lqae047.
9. Philpott M, Watson J, Thakurta A, Brown Jr T, Brown Sr T, Oppermann U, et al. Nanopore sequencing of single-cell transcriptomes with scCOLOR-seq. Nature biotechnology. 2021;39(12):1517-20.
10. PacificBiosciences. Application Note: Kinnex single-cell RNA kit for single-cell isoform sequencing. 2024.
11. XGenomics. Alternative transcript isoform detection with single cell and spatial resolution. 2022.
12. OxfordNanoporeTechnologies. Nanopore sequencing accuracy 2024. Available from: https://nanoporetech.com/platform/accuracy.
13. ParseBiosciences. Technical note: Preparing KINNEX™ single-cell libraries with Parse Biosciences EVERCODE™ WT kits. 2024.
14. ParseBiosciences. Multi-Omics Approach for Near Full Length Human iPSC Transcriptomes in Cardiomyocyte Models 2024. Available from: https://www.parsebiosciences.com/customer-datasets/multi-omics-approach-for-near-full-length-human-ipsc-transcriptomes-in-cardiomyocyte-models/.
15. Hagemann-Jensen M, Ziegenhain C, Sandberg R. Scalable single-cell RNA sequencing from full transcripts with Smart-seq3xpress. Nature Biotechnology. 2022;40(10):1452-7.
16. Hahaut V, Pavlinic D, Carbone W, Schuierer S, Balmer P, Quinodoz M, et al. Fast and highly sensitive full-length single-cell RNA sequencing using FLASH-seq. Nature biotechnology. 2022;40(10):1447-51.
17. Al'Khafaji AM, Smith JT, Garimella KV, Babadi M, Popic V, Sade-Feldman M, et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. Nature Biotechnology. 2024;42(4):582-6.
18. Created in BioRender. D., K. (2025) https://BioRender.com/v34e602.
19. ONT. Single-cell transcriptomics with 5' cDNA prepared using 10X Genomics on PromethION (SQK-LSK114). 2024.
20. Shiau C-K, Lu L, Kieser R, Fukumura K, Pan T, Lin H-Y, et al. High throughput single cell long-read sequencing analyses of same-cell genotypes and phenotypes in human tumors. Nature Communications. 2023;14(1):4124.
21. You Y, Prawer YDJ, De Paoli-Iseppi R, Hunt CPJ, Parish CL, Shim H, et al. Identification of cell barcodes from long-read single-cell RNA-seq with BLAZE. Genome Biology. 2023;24(1):66.
22. OxfordNanoporeTechnologies. wf-single-cell. 2024.

23. Zajac N, Zhang Q, Bratus-Neuschwander A, Qi W, Bolck HA, Karakulak T, et al. Comparison of Single-cell Long-read and Short-read Transcriptome Sequencing of Patient-derived Organoid Cells of ccRCC: Quality Evaluation of the MAS-ISO-seq Approach. bioRxiv. 2024:2024.03.14.584953.
24. Prjibelski AD, Mikheenko A, Joglekar A, Smetanin A, Jarroux J, Lapidus AL, et al. Accurate isoform discovery with IsoQuant using long reads. Nature Biotechnology. 2023;41(7):915-8.
25. Deng E, Shen Q, Zhang J, Fang Y, Chang L, Luo G, et al. Systematic evaluation of single-cell RNA-seq analyses performance based on long-read sequencing platforms. Journal of Advanced Research. 2024.
26. Pardo-Palacios FJ, Arzalluz-Luque A, Kondratova L, Salguero P, Mestre-Tomás J, Amorín R, et al. SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms. Nature Methods. 2024;21(5):793-7.
27. De Rijk P, Watzeels T, Küçükali F, Van Dongen J, Faura J, Willems P, et al. Scywalker: scalable end-to-end data analysis workflow for long-read single-cell transcriptome sequencing. Bioinformatics. 2024;40(9).
28. Westoby J, Artemov P, Hemberg M, Ferguson-Smith A. Obstacles to detecting isoforms using full-length scRNA-seq data. Genome Biology. 2020;21(1):74.
29. Wu S, Schmitz U. Single-cell and long-read sequencing to enhance modelling of splicing and cell-fate determination. Computational and Structural Biotechnology Journal. 2023;21:2373-80.
30. Dominguez D, Tsai Y-H, Weatheritt R, Wang Y, Blencowe BJ, Wang Z. An extensive program of periodic alternative splicing linked to cell cycle progression. Elife. 2016;5:e10288.
31. Petasny M, Bentata M, Pawellek A, Baker M, Kay G, Salton M. Splicing to keep cycling: the importance of pre-mRNA splicing during the cell cycle. Trends in Genetics. 2021;37(3):266-78.
32. Verwilt J, Mestdagh P, Vandesompele J. Artifacts and biases of the reverse transcription reaction in RNA sequencing. RNA. 2023;29(7):889-97.
33. Nam DK, Lee S, Zhou G, Cao X, Wang C, Clark T, et al. Oligo (dT) primer generates a high frequency of truncated cDNAs through internal poly (A) priming during reverse transcription. Proceedings of the National Academy of Sciences. 2002;99(9):6152-6.
34. Balázs Z, Tombácz D, Csabai Z, Moldován N, Snyder M, Boldogkői Z. Template-switching artifacts resemble alternative polyadenylation. BMC Genomics. 2019;20(1):824.
35. Schulz L, Torres-Diz M, Cortés-López M, Hayer KE, Asnani M, Tasian SK, et al. Direct long-read RNA sequencing identifies a subset of questionable exitrons likely arising from reverse transcription artifacts. Genome Biology. 2021;22(1):190.
36. Ramsköld D, Hendriks G-J, Larsson AJM, Mayr JV, Ziegenhain C, Hagemann-Jensen M, et al. Single-cell new RNA sequencing reveals principles of transcription at the resolution of individual bursts. Nature Cell Biology. 2024;26(10):1725-33.
37. Hagemann-Jensen M. Protocols.io Smart-seq3xpress V.2 Protocols.io; 2022.

38. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics. 2023;39(5):btad311.
39. Parekh S, Ziegenhain C, Vieth B, Enard W, Hellmann I. zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. GigaScience. 2018;7(6):giy059.
40. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology. 2020;38(3):276-8.
41. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nature Communications. 2017;8(1):14049.
42. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology. 2018;36(5):411-20.
43. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87.e29.
44. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nature Biotechnology. 2024;42(2):293-304.
45. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nature Biotechnology. 2015;33(5):495-502.
46. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888-902.e21.
47. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21(16):3439-40.
48. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature Protocols. 2009;4(8):1184-91.
49. Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome research. 2015;25(12):1860-72.
50. Martin Morgan HP, Valerie Obenchain, Nathaniel Hayden. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. 2023.
51. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS computational biology. 2013;9(8):e1003118.
52. Chen H. VennDiagram: Generate High-Resolution Venn and Euler Plots. 2022.
53. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. 2023.
54. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2016.
55. Hahne F, Ivanek R. Visualizing genomic data using Gviz and bioconductor. Statistical genomics: methods and protocols. 2016:335-51.
56. Gehlenborg N. UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets. 2019.

57. OxfordNanoporeTechnologies". pychopper. 2024.
58. Larsson A. stitcher.py. 2020.
59. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome research. 2017;27(3):491-9.

Figures

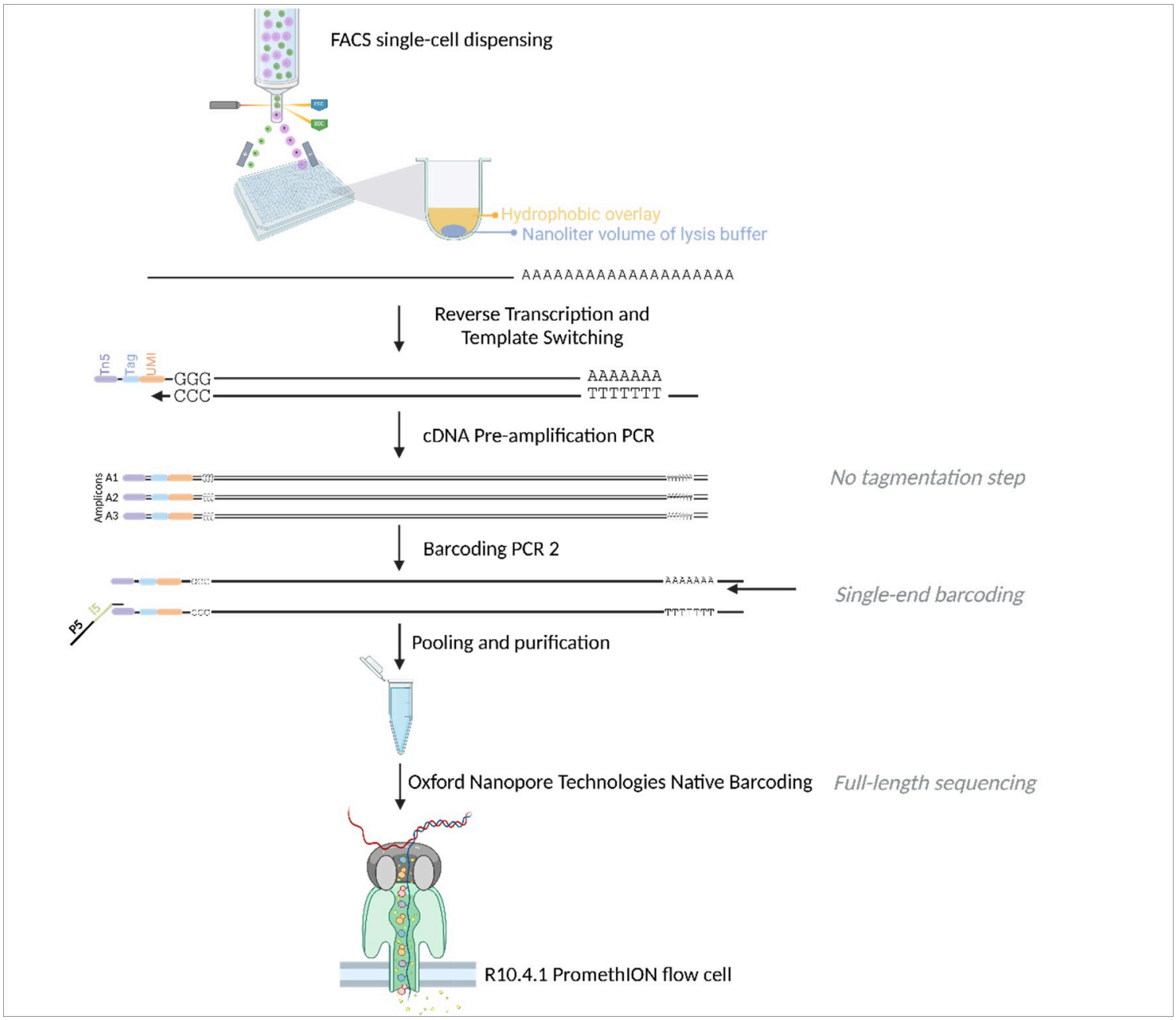

Figure 1

Overview of the scLIS-seq workflow. Single cells are isolated using FACS into a 384-well PCR-plate. The cells are lysed to release the RNA, subsequently followed by reverse transcription, template switching, and cDNA pre-amplification. The cDNA is 1:10 diluted, and 1 µL of the diluted cDNA is transferred to the

corresponding well of a new plate. Finally, this cDNA is further amplified and barcoded in PCR 2, followed by pooling, purification, and ONT library preparation. The main differences compared to SS3X are shown on the right.

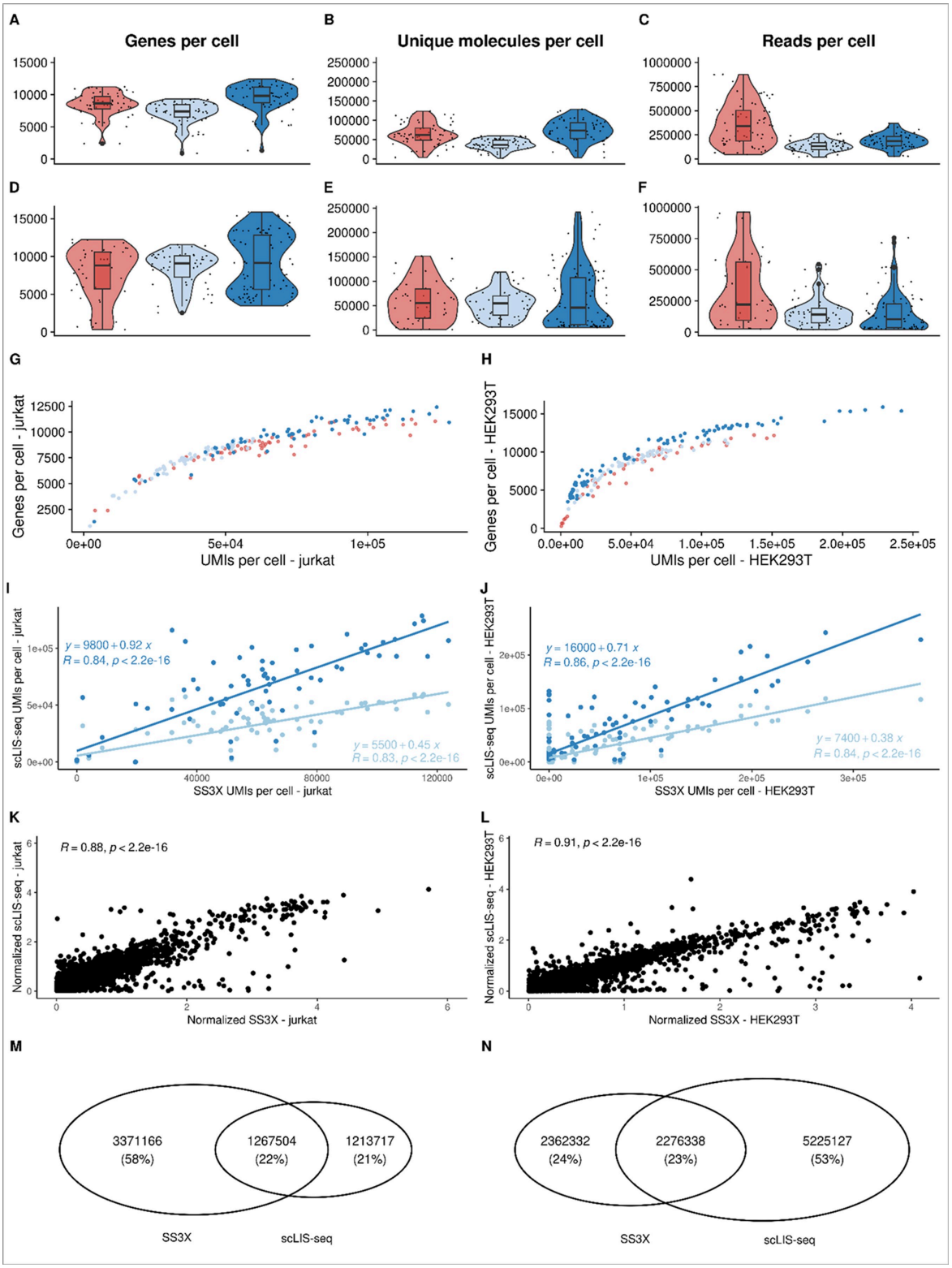

Figure 2

Performance of scLIS-seq compared to SS3X. Legend: Red: regular SS3X; light blue: scLIS-seq processed by wf-single-cell, dark blue: scLIS-seq processed by scywalker. **A, B, C,** Number of genes,

UMIs, and reads detected per Jurkat cell in regular SS3X using SRS and zUMIs processing, compared to scLIS-seq processed with the wf-single-cell pipeline and scywalker, respectively. **D, E, F,** Number of genes, UMIs, and reads detected per HEK293T cell in regular SS3X using SRS and zUMIs processing, compared to scLIS-seq processed with the wf-single-cell pipeline and scywalker , respectively. **G, H,** Scatter plots illustrating the relationship between the genes detected per cell and UMI counts per cell across the different strategies for Jurkat and HEK293T, respectively. **I, J,** Pair-wise scatter plots illustrating the number of UMIs detected in SS3X and scLIS-seq, for Jurkat and HEK293T, respectively. **K,L,** Pair-wise scatter plot comparing normalized gene expression between SS3X and scLIS-seq for Jurkat and HEK293T, respectively, as analyzed by wf-single-cell. **M, N,** Venn-diagrams illustrating the overlap of chromosome – cell barcode -UMI combinations between SS3X and scLIS-seq for Jurkat, analyzed by wf-single-cell and scywalker, respectively.

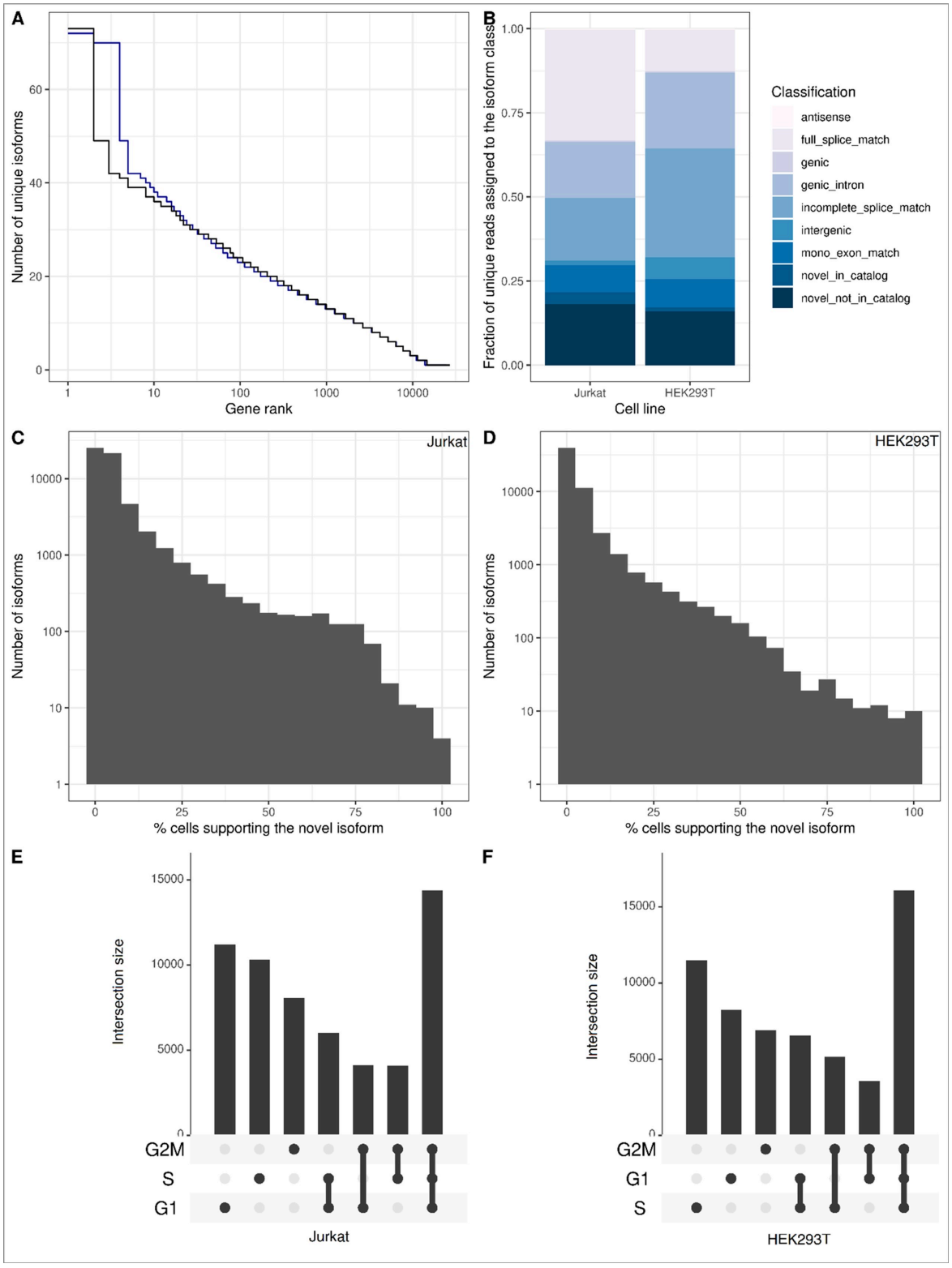

Figure 3

Isoform profiling using scLIS-seq. A. Gene rank plots illustrating the number of unique isoforms identified per gene for Jurkat (dark blue) and HEK293T (black). **B.**Classification of the identified isoforms in their SQANTI3 isoform class. The detailed definitions of each classification can be found in the SQANTI3 documentation. **C,D,**Percentage of cells supporting the detected novel isoform for Jurkat and

HEK293T, respectively. **E, F,** Number of novel isoforms classified according to their occurrence in cells within a specific cell cycle phase for Jurkat and HEK293T, respectively.

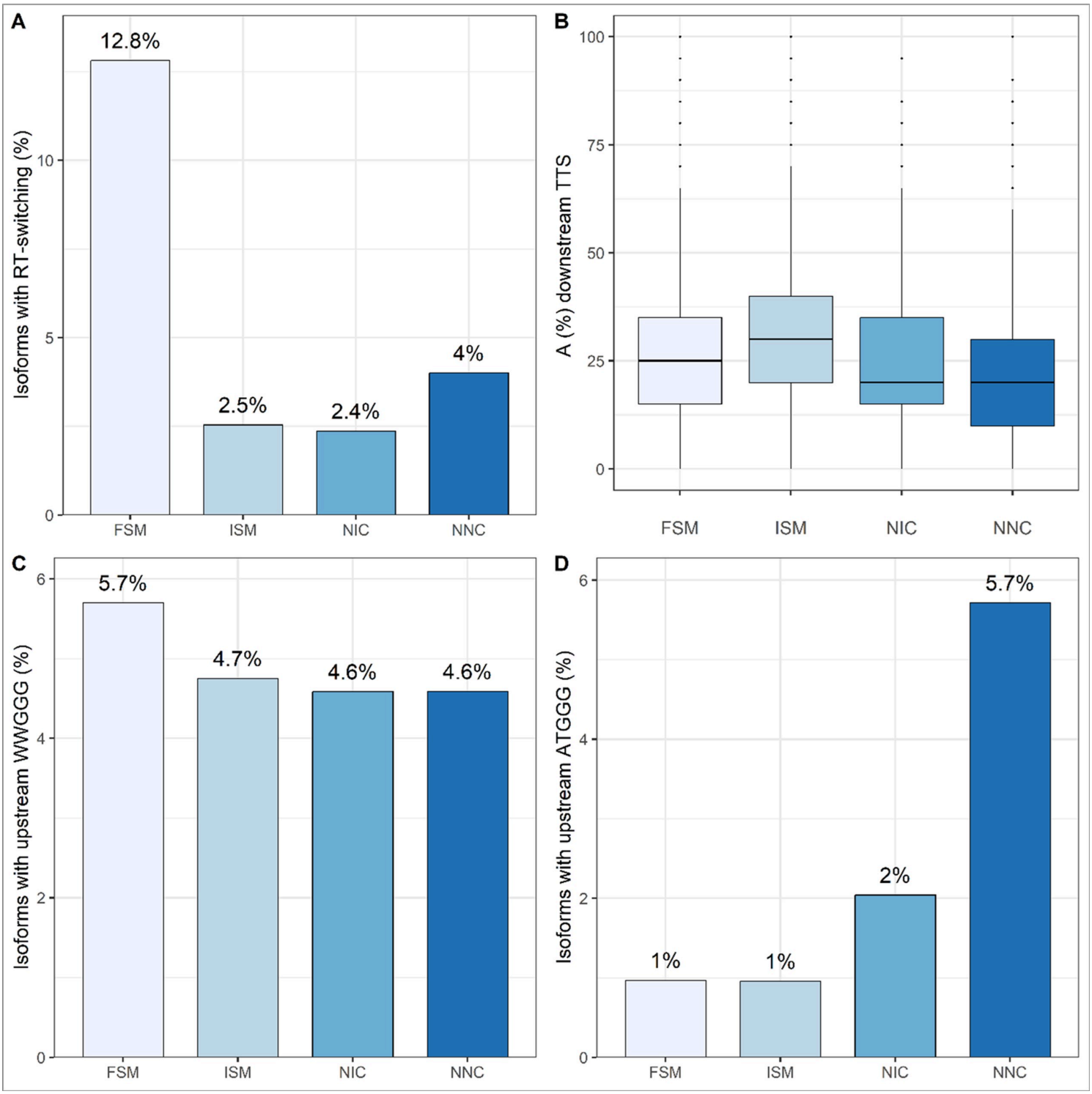

Figure 4

Overview of isoform classification based on the output of SQANTI3 for the HEK293T cells. A. Number of isoforms with suspected RT-switching event according to their respective isoform classification. **B.** Number of adenosine nucleotides in the 20-nucleotide genomic window downstream of the TTS to evaluate potential intra-priming events. **C.** Percentage of isoforms with suspected strand-invasion events

caused by TSO-mispriming for the scLIS-seq data. The calculation is based on the presence of the WWGGG TSO-motif in the 20-nucleotide genomic window upstream of the TSS. **D.** Percentage of isoforms with suspected strand-invasion events caused by TSO-mispriming identified from a subset of the reads extracted from the scNanoGPS A-375 cells. The calculation is based on the presence of the ATGGG TSO-motif in the 20-nucleotide genomic window upstream of the TSS. FSM: full-splice-match; ISM: incomplete-splice-match; NIC: novel-in-catalog; NNC: novel-not-in-catalog.

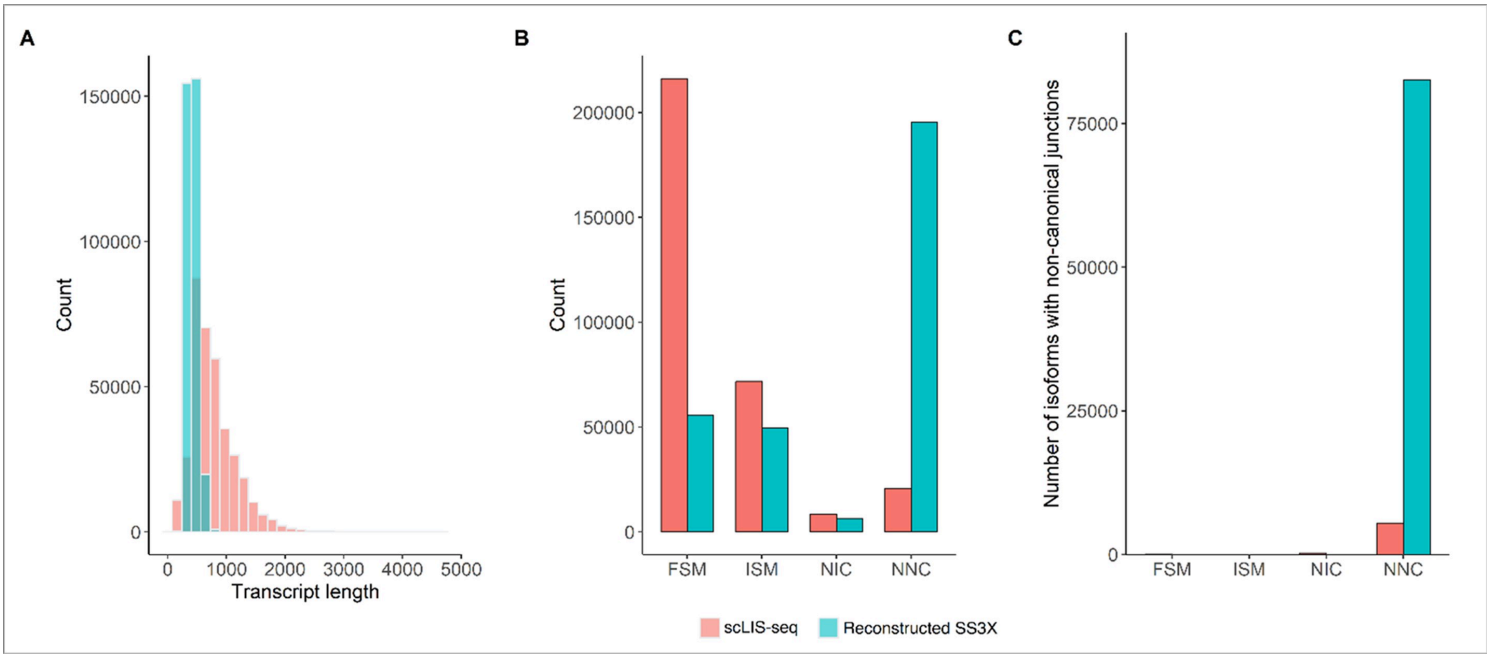

Figure 5

Direct comparison of the isoforms identified with scLIS-seq and reconstructed using SS3X for jurkat. **A.** Comparison of the isoform lengths between scLIS-seq and the isoforms reconstructed using UMI-stitching of the SR SS3X data. **B.** SQANTI3 classification of the identified isoforms for both strategies. **C.** Number of isoforms within each classification category containing non-canonical splice junctions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- scLISseqSupplementaryMaterials.pdf