Full-length bulk and single-cell RNA sequencing for cancer research

The ability to characterize bulk & single-cell transcriptomes at the isoform level is critical for cancer research. In bulk transcriptome data, short reads rely on computational assembly to infer transcripts, which is challenging for many alternatively spliced isoforms. In single-cell transcriptomics, short reads can only capture gene-level information, while other long-read technologies lack the accuracy needed for identification of unique molecular identifiers (UMI) and cell barcodes (CBC).

The Iso-Seq® method, full-length RNA sequencing using PacBio® HiFi sequencing, has been shown to advance cancer research by unambiguously characterizing isoforms at both the bulk and single-cell level. PacBio Kinnex™ RNA kits, based on the MAS-Seq method for concatenating amplicons such as cDNA into longer fragments, further increase throughput on HiFi long-read sequencers. HiFi reads generated from sequencing the concatenated molecules can then be bioinformatically broken up to retrieve the original cDNA sequences. Kinnex kits offer different concatenation factors based on target amplicon sizes to maximize throughput increase.

Here, we highlight publications using Iso-Seq in cancer research to demonstrate how bulk and single-cell Iso-Seq data detects fusion genes, identifies novel isoforms that might serve as neoepitopes for vaccine candidates, and traces clonal evolution.

102-326-634 REV01 MAR2025

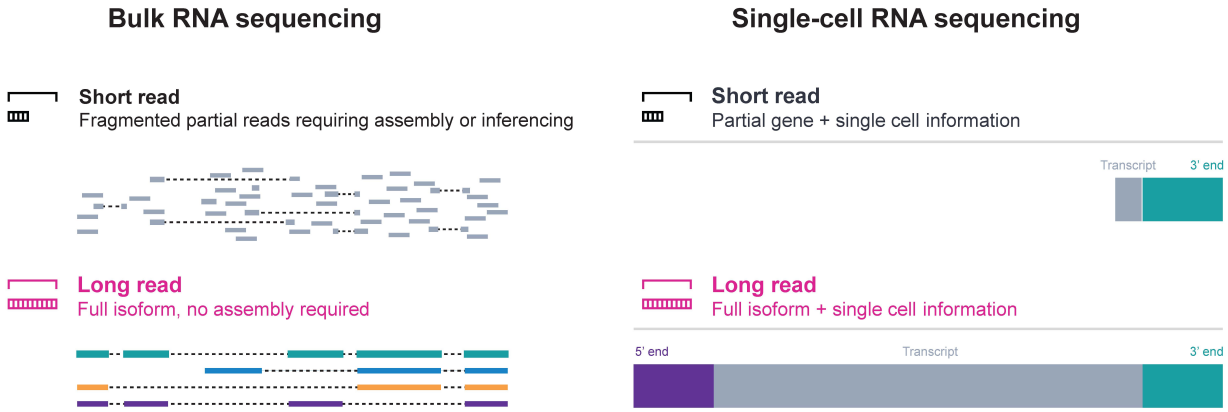

Figure 1. Comparison of long- vs. short-read RNA sequencing. With bulk transcriptomes, short reads required computational assembly or inference to identify the transcripts, whereas long reads capture the entire full-length cDNA. With single-cell transcriptomes, short reads are limited to the ends of a cDNA molecule and can only offer gene-level information. In contrast, long reads can provide isoform information at the single-cell level.

Fusion transcript detection

Short reads can readily identify fusion junctions, but due to their insufficient read length cannot fully resolve individual fusion transcript structures. For example, as demonstrated in Nattestad et al. (2018) using Iso-Seq, 3-hop fusion genes where the fusion spans more than 2 genes can be mis-identified by short reads as two separate events.

In another publication, Dondi et al. (2023) applied an early version of the Kinnex concatenation method to single-cell full-length RNA sequencing to ovarian cancer samples. In their Kinnex data, they identified a patient-specific *IGFBP2::TESPA1* fusion that was previously mis-classified as high *TESP1A1* expression in matched short-read data (Figure 2) that only captured the 3' fusion partner.

In Qin et al. (2024), the authors developed a new long-read fusion detection tool, CTAT-LR-fusion, and showed that it consistently detected all 16 control fusions in the SeraCare FusionStandard truth set with Kinnex single-cell data across replicates. Short reads, however, missed detecting 9 control fusions in some replicates.

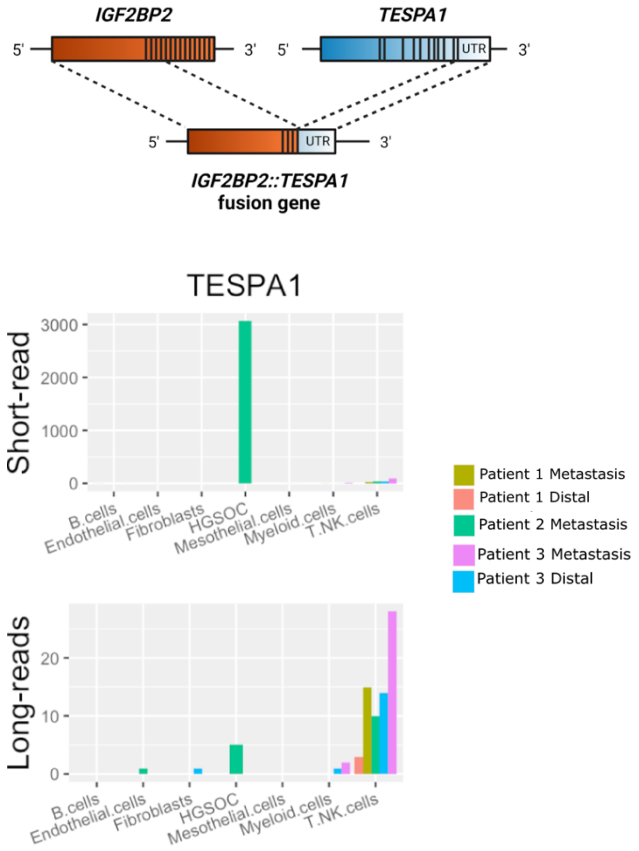

Figure 2. Long-read single-cell RNA sequencing identified a tumor-specific fusion gene *IGF2BP2::TESPA1* that was mis-classified by short read data as *TESPA1* overexpression. Figures available from Dondi et al. (2023) with no additional changes through its Creative Commons license.

Novel isoform discovery for translational research

Splicing dysregulation is a pervasive feature of most cancers, yet the general landscape of aberrant splicing is poorly understood. Improved mRNA profiling using the Iso-Seq method can identify new biomarkers and therapeutic targets (Anczukow et al., 2024).

Recent studies using Iso-Seq in both bulk and single-cell transcriptomes have revealed the extent to which novel isoforms are abundant in cancer. Veiga et al. (2022) profiled 26 breast cancer tumor samples and found 67% of the isoforms to be novel and were enriched in oncogenes and in many cases resulted in differential predicted protein structure or localization. The authors also found 35 survival-associated alternative splicing (AS) events that were enriched in its breast cancer dataset. Twenty-one of these 35 events (60%) were novel isoforms that were detected with Iso-Seq data but absent in GENCODE.

Wijeratne et al. (2024) applied an early version of the Kinnex concatenation method to tumor- and tumor-adjacent normal tissues in a pediatric glioma sample and identified differentially expressed novel isoforms that were difficult to resolve with short reads alone and could potentially serve as biomarkers.

In another study, Kohli et al. (2017) used targeted RNA sequencing to identify the entire exon composition of androgen receptor (AR) isoforms in castration-resistant prostate cancer (CRPC). The AR variant AR-V7 is known to promote prostate cancer resistance to AR-targeted therapies. Previous efforts using short read RNA-Seq had indicated additional AR-V species that are co-expressed with AR-V7 and may contribute to resistance. However, these studies could not resolve the specific co-expressed variants due to the presence of shared exons. Using Iso-Seq, they found that AR-V9 was frequently co-expressed with AR-V7 and may be a biomarker of increased drug resistance.

Huang et al. (2021) profiled ten gastric cancer cell lines with Iso-Seq and found that >66% of the isoforms detected were novel, predominantly driven by alternative first exon (AF) events. Notably, the *ARID1A* gene was found to have two new isoforms, both using a novel promoter that altered the protein coding sequence. Correlation with clinical datasets indicated these novel isoforms were significantly associated with poor progression-free survival in gastric cancer, whereas the known promoter usage was not (Figure 3).

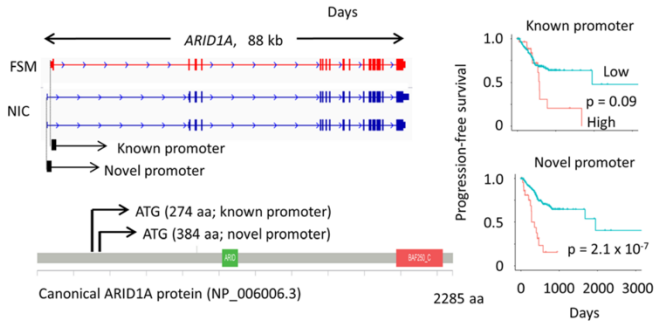

Figure 3. Novel isoforms using a novel promoter (NIC, blue) were associated with poor gastric cancer survival outcomes in Huang et al. (2021). Figure available with no additional changes through the article's Creative Commons license.

Finally, in Li et al. (2024), using single-cell RNA sequencing to study colorectal cancer (CRC) identified 394 dysregulated transcript structures in tumor epithelial cells. This study also identified both tumor- and isoform-specific RNA editing events. After filtering for tumor-specific novel isoforms with mass spectroscopy support, the authors developed an algorithm to rank these potential neoepitopes by HLA binding affinity. Using this approach, they found four novel tumor-specific isoforms with strong HLA binding affinity for a wide spectrum of CRC patients, which could potentially serve as cancer vaccine candidates.

Long-read fusion gene detection tools

Several fusion gene detection tools have been developed specifically for long-read RNA data, including CTAT-LR-fusion, pbfusion, PB_FLIP, and JAFFAL.

Tracing clonal evolution

Traditional short-read single-cell RNA-Seq only covers either the 5' or 3' end of a transcript and therefore is limited in its ability to determine a cell's genotype based on somatic tumor mutations. In contrast, single-cell Iso-Seq can sequence the entire cDNA from multiple single-cell platforms.

Black et al. (2024) applied the Kinnex single-cell RNA kit to single-cell cDNA generated by the 10x Chromium platform in chronic lymphocytic leukemia (CLL) samples harboring *BTK* resistance mutations, and they showed drastic improvement in genotyping capabilities. While Illumina short reads covered only 7.61% of the variants in ≥1% of the cells, Kinnex data covered 18.59% of these variants. This in turn helped refine the subclone structure of one of the samples, depicting two distinct mutated subclones that were previously mis-identified as a single subclone.

Similarly, Wedemeyer et al. (2024) found that when using the Kinnex single-cell RNA kit to sequence a patient-derived fibroblast with a known single nucleotide variant (*PIK3CA*:c.3139C>T), long reads covered the variant in 15.3% of the cells, as compared to only 1.2% of the cells using Illumina short reads. Additionally, by comparing the mutant cells against the wild-type cells, they identified distinct gene expressions and signaling pathways commonly associated with the epithelial-mesenchymal transition.

In summary, single-cell RNA sequencing increases variant detection and genotyping of single cells, which help improve tracing clonal evolution and somatic mosaicism in cancer research.

Conclusion

Long-read RNA sequencing (the Iso-Seq method) for bulk and single-cell transcriptomics can reveal important insights for cancer research not readily accessible using short reads. The publications highlighted here utilize the bulk and single-cell Iso-Seq methods, often combined with Kinnex kits for higher throughput, to detect novel isoforms, fusion transcripts, and neoepitopes, as well as to genotype single cells and trace clonal populations, yielding a more complete view of the cancer transcriptome.

Table 1. Summary of cancer research publications utilizing full-length RNA sequencing.

Study	Samples	Method	Key findings
Nattestad et al. (2018)	SKBR3 cell line	WGS, Iso-Seq	SVs, 2- and 3-hop fusion genes
Dondi et al. (2023)	Ovarian cancer	Single-cell Iso-Seq with Kinnex*	Cell-type specific tumor isoforms, patient-specific fusion gene
Qin et al. (2024)	SeraCare FusionStandard	Single-cell Iso-Seq with Kinnex*	CTAT-LR-fusion long-read fusion detection tool
Black et al. (2024)	Chronic lymphocytic leukemia (CLL) samples with BTK resistance	Single-cell Iso-Seq with Kinnex*	Improved single cell genotyping and clonal evolution tracing
Wedemeyer et al. (2024)	Patient-derived PIK3CA-altered fibroblast	Targeted single-cell Iso-Seq with Kinnex*	Improved single cell genotyping
Veiga et al. (2022)	Breast cancer cell lines and tumor samples	Iso-Seq	Tumor-associated novel isoforms
Kohli et al. (2017)	Prostate cancer	Targeted Iso-Seq	Co-expression of AR-V9 with AR-V7
Huang et al. (2022)	Gastric cancer	Iso-Seq	Novel isoforms with alternative promoters associated with survival outcome
Li et al. (2024)	Colorectal cancer	Single-cell Iso-Seq	Tumor-specific isoforms, neoepitopes
Wijeratne et al. (2024)	HBR_SIRV, pediatric glioma sample	Iso-Seq with Kinnex*	Differentially expressed novel isoforms

*Includes use of the original or modified versions of the MAS-Iso-Seq method, the MAS-Seq for 10x Single Cell 3' kit, and all Kinnex RNA kits.

References

Al'Khafaji, A. M., et al. (2023). High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nature Biotechnology*, 1-5. https://doi.org/10.1038/s41587-023-01815-7

Anczukow, O., et al. (2024). Steering research on mRNA splicing in cancer towards clinical translation. *Nature Reviews Cancer*, *24*(12), 887-905. https://doi.org/10.1038/s41568-024-00750-2

Black et al. (2024). Long-read single-cell RNA sequencing enables the study of cancer subclone-specific genotype and phenotype in chronic lymphocytic leukemia. *BioRxiv*. https://doi.org/10.1101/2024.03.15.585298

Dondi, A., Lischetti, U., et al. (2023). Detection of isoforms and genomic alterations by high-throughput full-length single-cell RNA sequencing in ovarian cancer. *Nature Communications*, *14*(1), 7780. https://doi.org/10.1038/s41467-023-43387-9

Huang, K. K., et al. (2021). Long-read transcriptome sequencing reveals abundant promoter diversity in distinct molecular subtypes of gastric cancer. *Genome Biology*, *22*, 1-24. https://doi.org/10.1186/s13059-021-02261-x

Kohli, M., et al. (2017). Androgen receptor variant AR-V9 is coexpressed with AR-V7 in prostate cancer metastases and predicts abiraterone resistance. *Clinical Cancer Research*, *23*(16), 4704-4715. https://doi.org/10.1158/1078-0432.CCR-17-0017

Li, Z., Zhang, B., Chan, J. J., et al. (2024). An isoform-resolution transcriptomic atlas of colorectal cancer from long-read single-cell sequencing. *Cell Genomics*, *4*(9). https://doi.org/10.1016/j.xgen.2024.100641

Nattestad, M., et al. (2018). Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. *Genome Research*, *28*(8), 1126-1135. https://doi.org/10.1101/gr.231100.117

Qin et al. (2024). CTAT-LR-fusion: accurate fusion transcript identification from long and short read isoform sequencing at bulk or single cell resolution. *BioRxiv*. https://doi.org/10.1101/2024.02.24.581862

Veiga, D. F., Nesta, et al. (2022). A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. *Science Advances*, *8*(3), eabg6711. https://doi.org/10.1126/sciadv.abg6711

Wedemeyer, M. A., et al. (2024). Defining the transcriptome of PIK3CA-altered cells in a human capillary malformation using single cell long-read sequencing. *Scientific Reports*, *14*(1), 25440. https://doi.org/10.1038/s41598-024-72167-8

Wijeratne, S., et al. (2024). Full-length isoform concatenation sequencing to resolve cancer transcriptome complexity. *BMC Genomics, 25(1),* 122. https://doi.org/10.1186/s12864-024-10021-x

Resources

Application note – Kinnex full-length RNA kit for isoform sequencing

Application note – Kinnex single-cell RNA kit for single-cell isoform sequencing

Webinar – The Full-Length Transcriptomic Atlas of Human Colorectal Cancer from Single-Cell Isoform Sequencing (speaker: Zhongxiao Li)

Webinar – The RNA isoform landscape of cancer (speaker: Olga Anczukow)

Webinar – Detecting cancer-related RNA dysregulation with long-read sequencing (speaker: Arthur Dondi)

Webinar – Understanding clonal evolution using game theory and single-cell long-read isoform analysis (speaker: Nathan Salomonis)