PacBio

Application note

Kinnex single-cell RNA kit for single-cell isoform sequencing

Introduction

Understanding cell heterogeneity at the isoform level is critical for both basic and disease research. Short reads can only capture gene-level information, while other long-read technologies lack the accuracy for accurate unique molecular identifiers (UMI) and cell barcode (CBC) identification. PacBio® HiFi reads sequence full-length RNA isoforms along with single-cell barcode and UMI information, revealing extraordinary insight into single-cell biology.

The *Kinnex™ single-cell RNA kit* takes as input single-cell cDNA and outputs a sequencing-ready library that results in a 16-fold throughput increase compared to regular single-cell Iso-Seq® libraries. Combined with isoform-aware single-cell analysis SMRT® Link software, PacBio offers cost-effective single-cell isoform sequencing that does not require orthogonal sequencing methods. The SMRT Link software supports bioinformatics analysis to produce an isoform-level single-cell data matrix compatible with tertiary analysis software.

102-326-549 REV06 AUG2024

PacBio

Single-cell RNA sequencing

Single-cell RNA sequencing (scRNA-seq) emerged to characterize gene expression differences between individual cells derived from a complex tissue, allowing a higher-resolution view of the transcriptome.

Most single-cell experiments are done with short reads, which only capture the ends of molecules due to fragmentation. Sequencing fragments limits expression information to the gene level, missing important isoform diversity that could be important for disease or biological function.

PacBio HiFi reads sequence full-length RNA isoforms along with single-cell barcode and UMI information (Figure 1), revealing isoform diversity at the single-cell level.

HiFi sequencing advantages for single-cell RNA sequencing

- Full-length isoform information
- Accurate cell barcode and UMI detection
- Variant detection

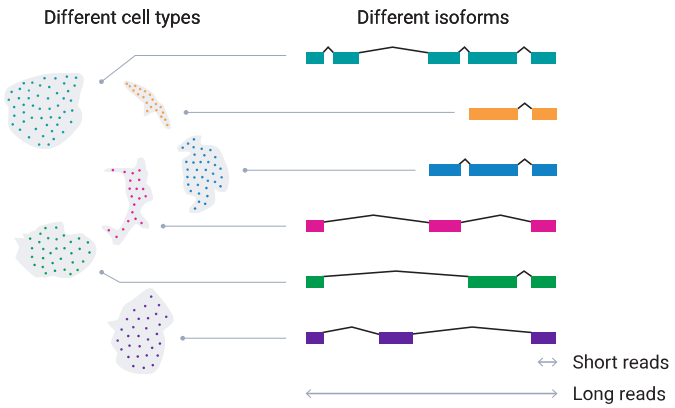

Figure 1. Single-cell isoform sequencing with PacBio long reads. Short reads only capture gene ends, missing isoform diversity. PacBio HiFi reads cover the entire isoform along with the single-cell barcode and UMI information with high accuracy.

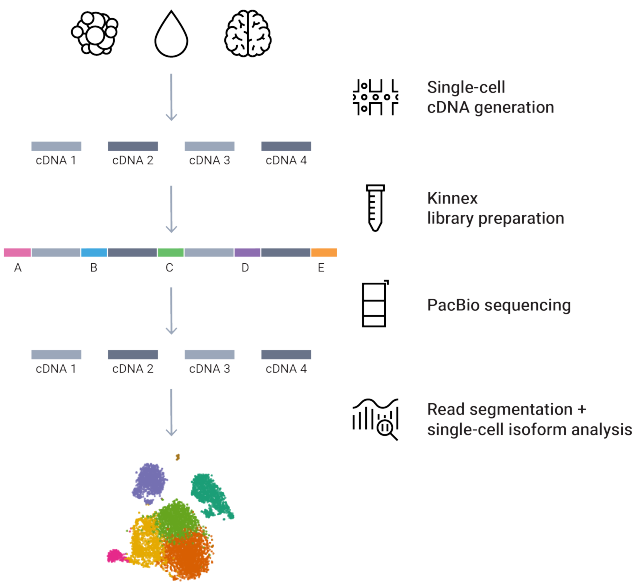

Figure 2. Kinnex for single-cell isoform sequencing. Single-cell cDNA molecules are concatenated into a larger insert library and sequenced, then processed using the SMRT Link software.

Kinnex for single-cell isoform sequencing

The Kinnex single-cell RNA kit utilizes the MAS-Seq method for throughput increase. MAS-Seq (Al'Khafaji et al., 2023) is a concatenation method for increasing throughput by joining cDNA molecules into longer concatenated fragments. HiFi reads generated from sequencing the concatenated molecules can then be bioinformatically broken up to retrieve the original cDNA sequences. The result is higher throughput and reduced sequencing needs for cost-effective single-cell isoform sequencing. Traditionally, orthogonal short-read scRNA-Seq is used to supplement the lower throughput of regular single-cell Iso-Seq method — with Kinnex, no orthogonal sequencing data are required.

The PacBio Single-cell Iso-Seq workflow processes the full-length cDNA sequences to classify them against a reference annotation (e.g., GENCODE) to identify novel genes and isoforms. The output consists of gene- and isoform-level count matrices that are compatible with tertiary analysis software.

The Kinnex kit is compatible with cDNA generated using the *10x Chromium Next GEM Single Cell 3' kit* (v3.1) and *Single Cell 5' kit* (v2) and is intended for use on a 3,000 to 10,000 cell library with 15-75 ng of cDNA as input.

Kinnex single-cell library workflow

The Kinnex library workflow begins with single-cell cDNA and produces a Kinnex library that is ready for sequencing.

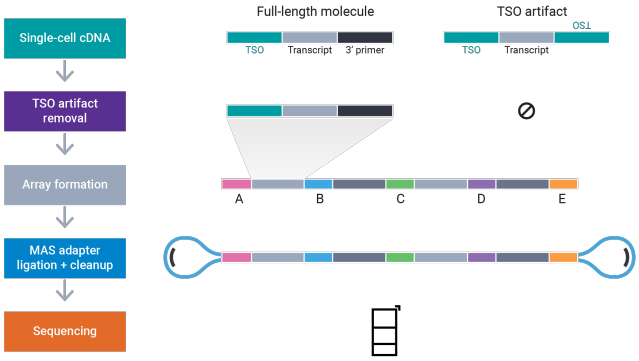

Figure 3. Kinnex single-cell library workflow.

Single-cell cDNA molecules are first removed of TSO (template-switching oligos) artifacts, then concatenated to form ordered arrays. Kinnex adapter ligation ensures full arrays are enriched, which are then sequenced on PacBio long-read sequencers (Figure 3). Additionally, the barcoded Kinnex adapters allow multiple Kinnex libraries to sequence on one SMRT® Cell.

With proper full array formation and adequate sequencing, one SMRT Cell on the Sequel® II/IIe and Revio™ systems are expected to achieve 30–40 million and 80–100 million cDNA sequences, respectively (Table 1).

Table 1. Target Kinnex single-cell library performance

Metric	Performance
Sample preparation time	2 days
Expected library size	11,000–14,000 bp
Target P1 loading	60–80%
Expected HiFi yield per SMRT Cell	2–2.5 million HiFi reads (Sequel II/IIe) 5–6 million HiFi reads (Revio)
Expected full array %	85–92%
Expected read yield per SMRT Cell	30–40 million reads (Sequel II/IIe) 80–100 million reads (Revio)

Kinnex single-cell bioinformatics workflow

The SMRT Link *Read Segmentation and Single-cell Iso-Seq* workflow processes the HiFi reads generated from the Kinnex library to produce gene- and isoform-level count matrices that are compatible with tertiary single-cell analysis tools (Figure 4).

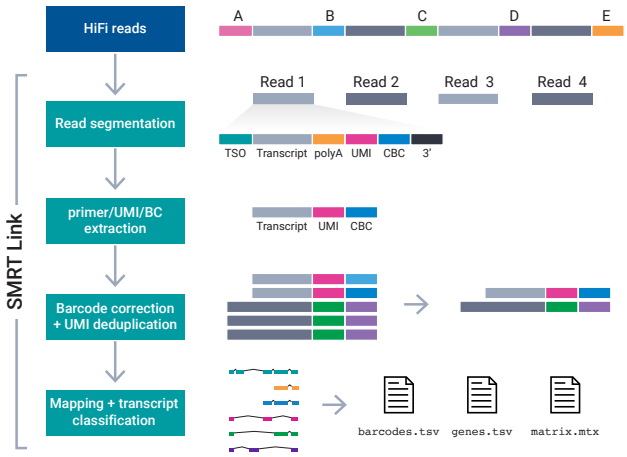

Figure 4. Kinnex analysis using *Read Segmentation and Single-cell Iso-Seq* workflow which is available in SMRT Link v11.1 and up.

Read segmentation

HiFi reads are segmented into individual segmented reads (*S-reads*) that represent the original cDNA sequences.

Primer/UMI/BC extraction

Primers and polyA tails are removed, but also used to orient the read into 5′ – 3′ orientation. Single-cell barcode and UMI information are extracted.

Barcode correction + UMI deduplication

Cell barcodes are corrected given an expected barcode list. Real cells — cell barcodes that represent encapsulated single cells (as opposed to ambient RNA) are also identified at this step. Reads are then deduplicated based on cell barcodes and UMIs.

Mapping and transcript classification

Deduplicated reads are mapped to the reference genome and classified against a transcript annotation (e.g., GENCODE). Finally, a gene- and isoform-level single-cell matrix is output for tertiary analysis.

Metric	Revio PBMC (3')	Revio PBMC (5')	Revio HG002 (5')
HiFi reads	8,889,073	6,104,086	6,460,787
Segmented reads (S-reads)	141,045,827	91,323,803	98,964,035
Mean S-read length	860 bp	980 bp	1,039 bp
S-reads with valid barcodes	134,118,572 (95%)	88,711,954 (97%)	96,607,591 (98%)
Deduplicated reads	81,678,354	44,323,585	82,910,986
Estimated number of cells	12,840	13,984	9,738
Reads in cells	84.7%	64.6%	91.3%
Mean reads per cell	8,852	4,146	9,121
Median UMIs per cell	4,884	2,498	7,449
Median genes per cell	912	697	1,706
Median transcripts per cell	1,060	771	2,042

Table 2. Read, cell, and transcript statistics of Kinnex single-cell library runs on Sequel II/IIe and Revio systems. Each library was run on one SMRT Cell. Data can be downloaded here.

Kinnex example: HG002 and PBMC dataset

PBMC cDNA generated using the *10x Chromium Next GEM Single Cell 5' and 3' kit* was made into Kinnex libraries and sequenced with one SMRT Cell each and analyzed using the *Read Segmentation and Single-Cell Iso-Seq* workflow in SMRT Link v13.1 (Table 2). The output results were then processed with tertiary tools to identify cell types (Figure 5) for the PBMC dataset. The HG002 cell line data was generated with a matching Illumina dataset showing high concordance in UMI counts for the recovered cell barcodes (Figure 6).

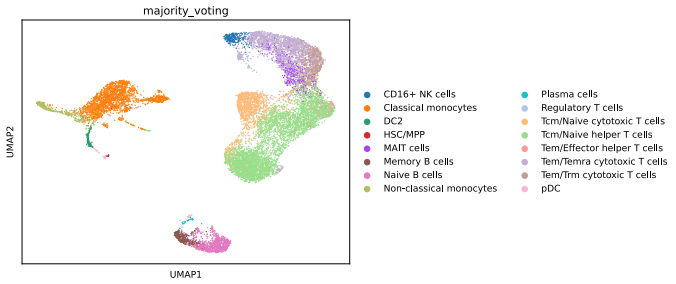

Figure 5. UMAP cell clustering of Revio-PBMC (10x 5' library) using CellTypist after SMRT Link analysis.

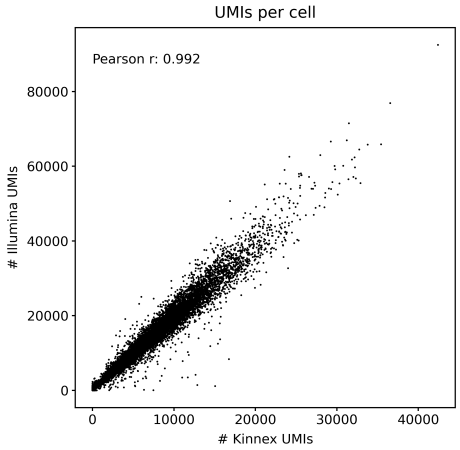

Figure 6. High UMI count concordance for recovered cells from matching Kinnex and Illumina HG002 10x 5' library. Pearson correlation: 0.992.

Kinnex for single-cell isoform sequencing: summary

The *Kinnex single-cell RNA kit* offers an end-to-end solution for single-cell RNA isoform sequencing from sample preparation to bioinformatics analysis.

- Supports cDNA from *10x Chromium Next GEM Single Cell 3' kit* (v3.1, v4) and *Single Cell 5' kit* (v2, v3) and Parse Biosciences Evercode WT kits
- Supports multiplexing with barcoded Kinnex adapters, up to 4-plex
- 15–75 ng input cDNA
- Target 3,000 to 10,000 cell library
- 16-fold throughput increase compared to non-Kinnex methods
- No orthogonal sequencing data required

For more information, visit https://pacb.com/kinnex

References

Al'Khafaji, A. M., et al. (2023). High-throughput RNA isoform sequencing using programmed cDNA concatenation. Nature Biotechnology, 1-5.