MAS-Seq: Towards isoform resolution single-cell transcriptomics using highly accurate long-read sequencing

Greg Young, Elizabeth Tseng, Roger Volden, Jason Underwood, Jessica Mattick, Ting Hon, Suzanne Dee

PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025, USA

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 unprecedented insight into single-cell biology.

The **MAS-Seq for *10x Single Cell* 3' kit** takes as input single-cell cDNA and outputs a sequencing-ready library that results in a 16-fold throughput increase. Combined with an 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 MAS-Seq bioinformatics analysis to produce isoform-level single-cell data matrix compatible with tertiary analysis software.

Single-cell isoform 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.

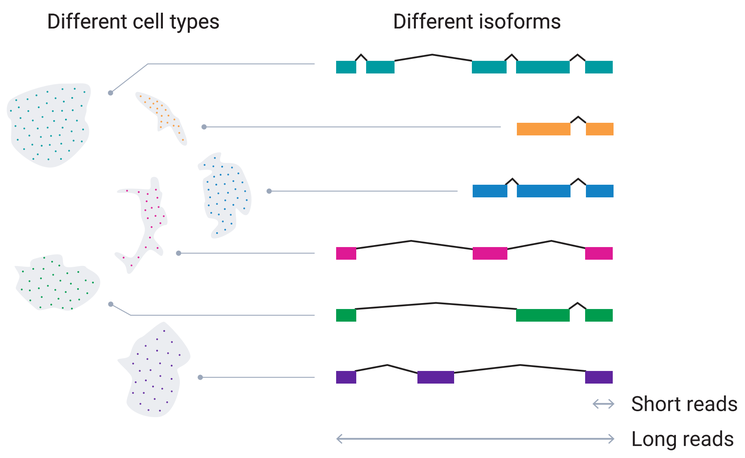

Figure 1. Single-cell isoform sequencing with PacBio long reads. Short reads only capture gene ends, missing isoform diversity. PacBio HiFi reads cover the entire isoform along with the single-cell barcode and UMI information with high accuracy.

Revio

MAS-Seq

MAS-Seq for single-cell isoform sequencing

The *multiplexed arrays sequencing* (MAS-Seq) method increases 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.

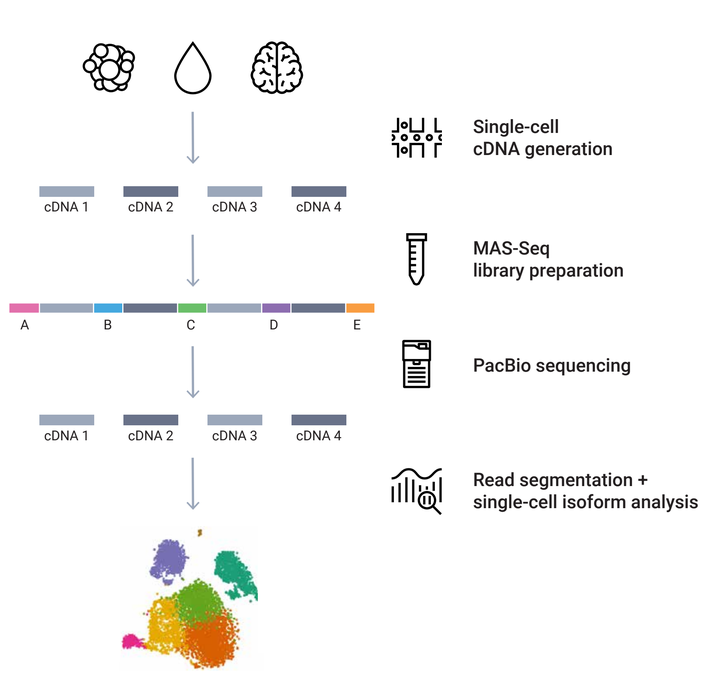

Figure 2. MAS-Seq for single-cell isoform sequencing. Single-cell cDNA molecules are concatenated into a larger insert library and sequenced, then processed using the PacBio software.

MAS-Seq library workflow

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

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

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

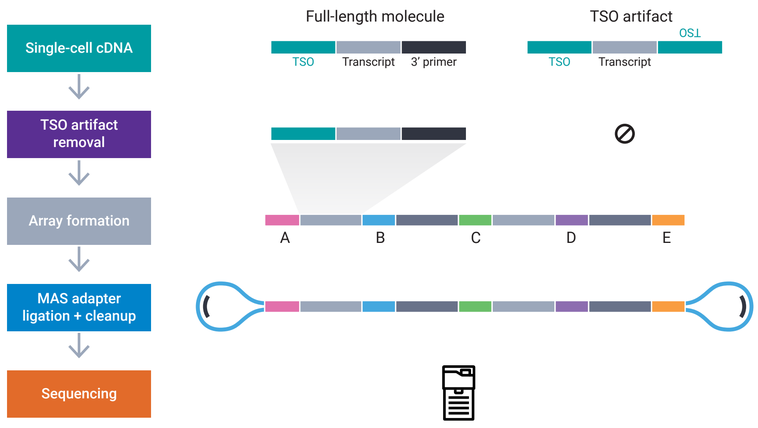

Figure 3. MAS-Seq library workflow.

Table 1. Target MAS-Seq library performance

Metric	Performance
Sample preparation time	2 days
Expected library size	11,000–14,000 bp
Target P1 loading	60–80%
Expected full array %	85–92%
Expected read yield	Sequel II/IIe system: 40 million reads Revio system: 100 million reads

MAS-Seq bioinformatics workflow

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

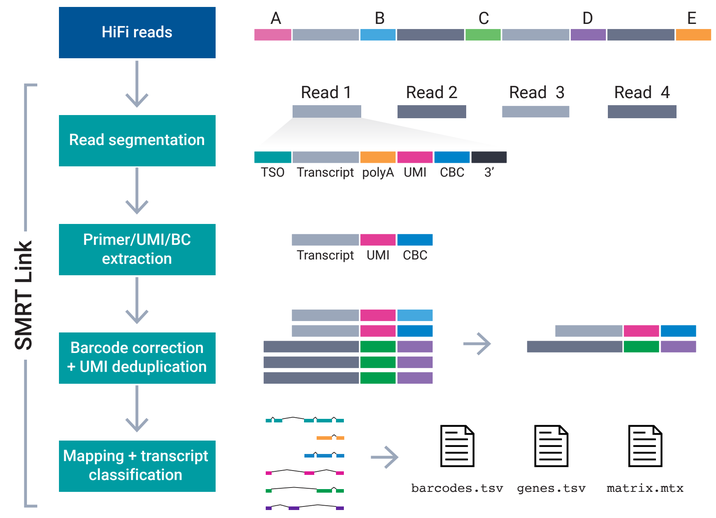

Figure 4. MAS-Seq analysis using read segmentation and single-cell Iso-Seq workflow.

Example: PBMC MAS-Seq dataset on Sequel IIe and Revio systems

Two sets of PBMC cDNA generated using the 10x *Chromium Single Cell* 3' kit were made into MAS-Seq libraries and sequenced on one SMRT Cell each and analyzed using the *Read Segmentation & Single-Cell Iso-Seq* workflow in SMRT Link v11.1 (Table 2). The output results were then processed with tertiary tools to identify cell types (Figure 5).

Table 2. Read, cell, and transcript statistics of two sample PBMC MAS-Seq library runs

Metric	Sequel IIe sample A	Revio sample B
HiFi reads	2,622,891	7,261,030
Segmented reads (S-reads)	40,131,832	110,127,016
S-read length distribution	300–2,500 bp	300–2,500 bp
S-reads with valid barcodes	37,634,585 (94%)	105,309,352 (96%)
Deduplicated reads	23,883,685	60,421,293
Est. number of cells	3,966	8,822
Reads in cells	91.6%	94.99%
Mean reads per cell	8,708	11,352
Median UMIs per cell	4,821	5,861
Median genes per cell	704	938
Median transcripts per cell	818	1,103

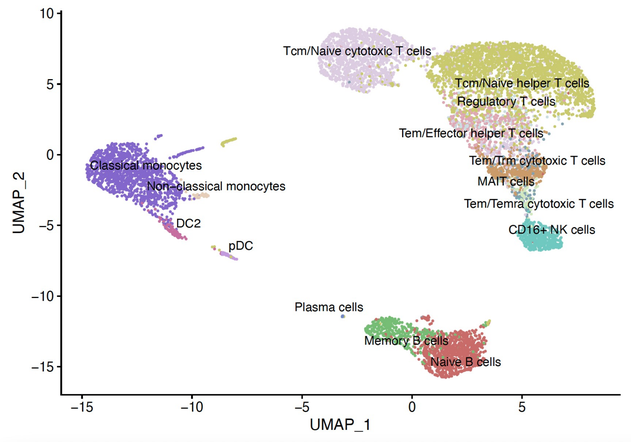

Figure 5. UMAP cell clustering of Revio PBMC MAS-Seq data using CellTypist after SMRT Link v11.1 analysis.