Full-Length cDNA Sequencing on the PacBio Sequel Platform

Ting Hon, Elizabeth Tseng, Aparna Vedula, and Tyson A. Clark
PacBio, 1380 Willow Road, Menlo Park, CA 94025

Abstract

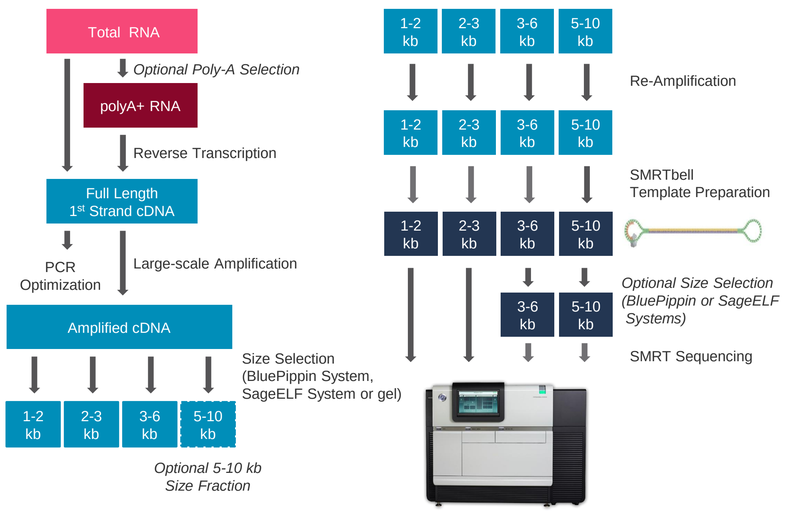
The protein coding potential of most plant and animal genomes is dramatically increased via alternative splicing. Identification and annotation of expressed mRNA isoforms is critical to the understanding of these complex organisms. While microarrays and other NGS-based methods have become useful for studying transcriptomes, these technologies yield short, fragmented transcripts that remain a challenge for accurate, complete reconstruction of splice variants.

The Iso-Seq protocol developed at PacBio offers the only solution for direct sequencing of full-length, single-molecule cDNA sequences to survey transcriptome isoform diversity useful for gene discovery and annotation. Knowledge of the complete isoform repertoire is also key for accurate quantification of isoform abundance. As most transcripts range from 1 – 10 kb, fully intact RNA molecules can be sequenced using SMRT Sequencing without requiring fragmentation or post-sequencing assembly. The Sequel System has improved throughput thereby increasing the number of full-length transcripts per SMRT Cell. Furthermore, loading enhancements on the Sequel instrument have decreased the need for size fractionation steps. We have optimized the Iso-Seq library preparation process for use on the Sequel platform.

Here, we demonstrate the capabilities of the Iso-Seq method on the Sequel System using cDNAs from the maize (Zea mays) inbred line B73. Full-length cDNA from six diverse tissues were barcoded, pooled, and sequenced on the PacBio Sequel System using a combination of size-selected and non-size-selected SMRTbell libraries. The results highlight the value of full-length transcripts for genome annotations and analysis of alternative splicing.

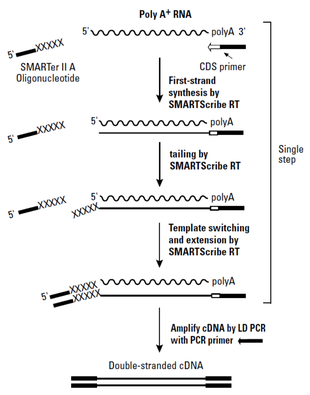
Iso-Seq Sample Preparation Methods

Iso-Seq Sample Preparation Workflow on the PacBio RS II

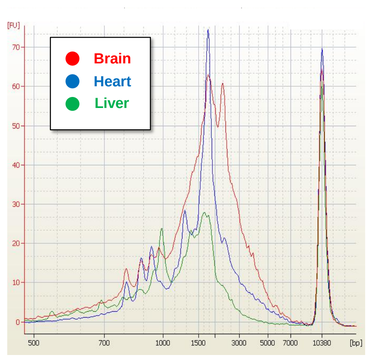


RNA is converted into first strand cDNA using the Clontech SMARTer PCR cDNA Synthesis Kit followed by universal amplification. Amplified cDNA is size fractionated and converted into SMRTbell templates for sequencing on the PacBio RS II.

Clontech SMARTer PCR cDNA Synthesis Kit

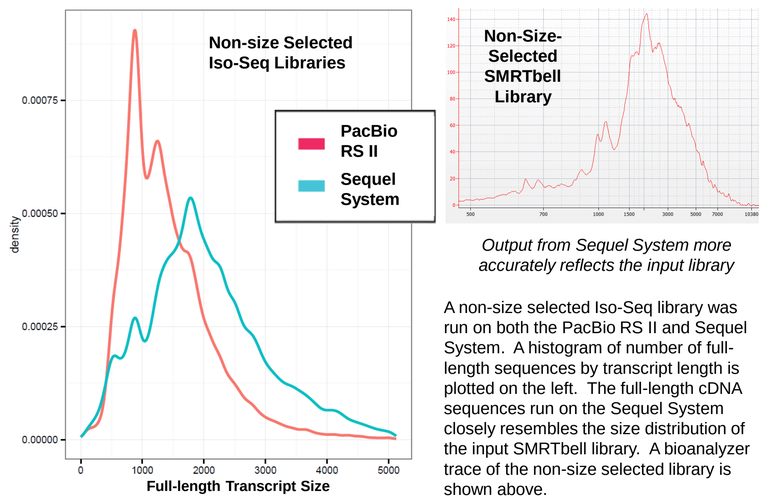


Size Distribution of Amplified cDNA from Multiple Tissues



Optimizing the Iso-Seq Application on the Sequel Platform

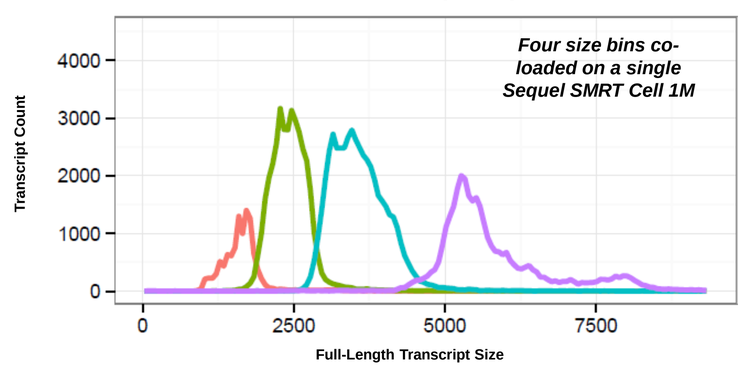
Magbead-loaded Samples on Sequel System Have Decreased Loading Bias



Output from Sequel System more accurately reflects the input library

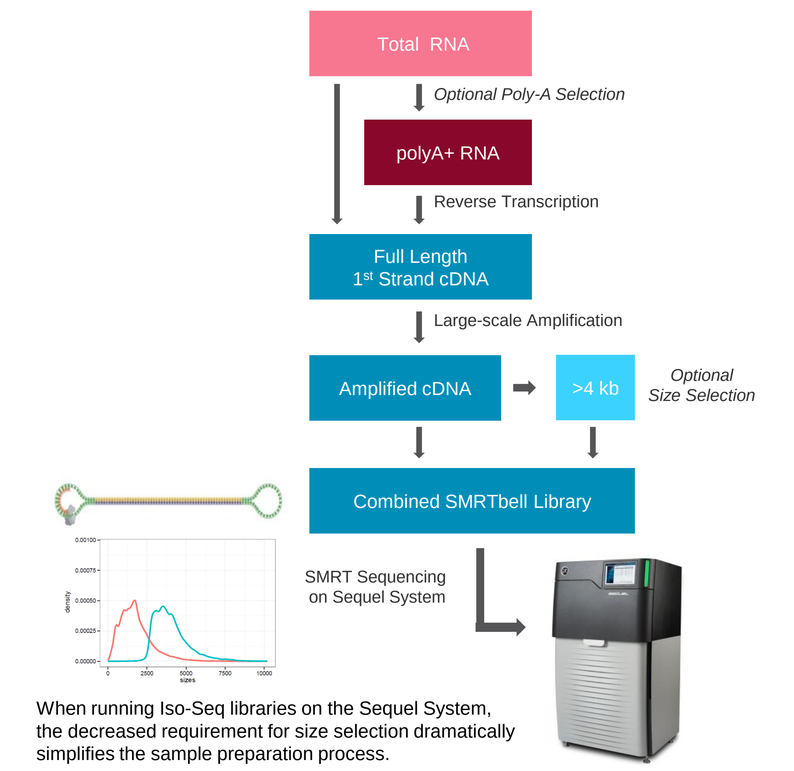
A non-size selected Iso-Seq library was run on both the PacBio RS II and Sequel System. A histogram of number of full-length sequences by transcript length is plotted on the left. The full-length cDNA sequences run on the Sequel System closely resembles the size distribution of the input SMRTbell library. A bioanalyzer trace of the non-size selected library is shown above.

Multiple Size-selected Libraries Can Be Co-Loaded on the Sequel System



Size-selected Iso-Seq libraries were combined and run together on a single Sequel SMRT Cell 1M. Amplified cDNA had been size fractionated into four size bins (1-2 kb, 2-3 kb, 3-6 kb, and 5-10 kb) using the Sage Science BluePippin system. SMRTbell libraries were made from each of the four size bins using separate barcoded hairpin adapters. A histogram of the full-length transcript lengths from each of the four size bins is shown above.

Streamlined Workflow for Iso-Seq Sample Preparation on the Sequel System



When running Iso-Seq libraries on the Sequel System, the decreased requirement for size selection dramatically simplifies the sample preparation process.

Barcoding cDNA Libraries

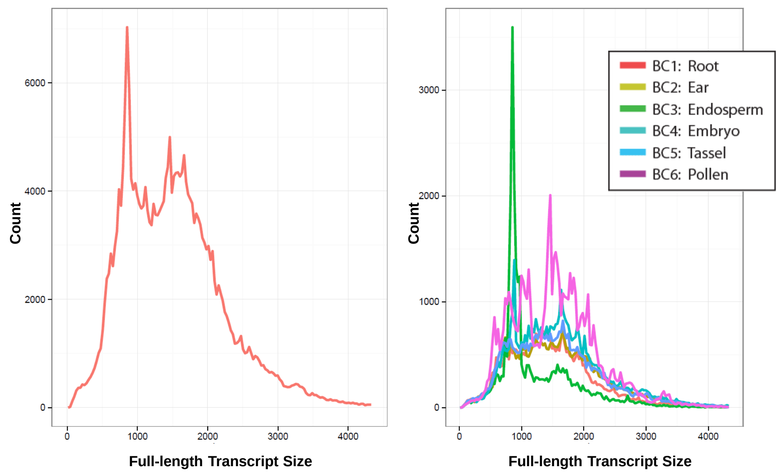
Barcoding During Reverse Transcription Step

SMARTer_dT_BC1	AAGCAGTGGTATCAACGCAGAGTACtcagacgatgcgtcatTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
SMARTer_dT_BC2	AAGCAGTGGTATCAACGCAGAGTACctatacatgactctgcTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
SMARTer_dT_BC3	AAGCAGTGGTATCAACGCAGAGTACtactagagtagcactcTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
SMARTer_dT_BC4	AAGCAGTGGTATCAACGCAGAGTACtgtgtatcagtacatgTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
SMARTer_dT_BC5	AAGCAGTGGTATCAACGCAGAGTACgatctctactatatgcTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
SMARTer_dT_BC6	AAGCAGTGGTATCAACGCAGAGTACacagtctatactgctgTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN

Primers containing 16-mer barcodes are used to prime first strand cDNA synthesis. RNA samples are reverse transcribed individually, then pooled prior to size fractionation.

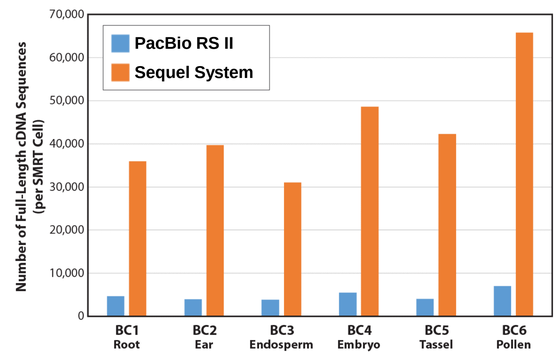
Analyzing the Maize Transcriptome

Barcoded Iso-Seq Libraries Generated from Six Diverse Maize (B73) Tissues and Run on the Sequel System



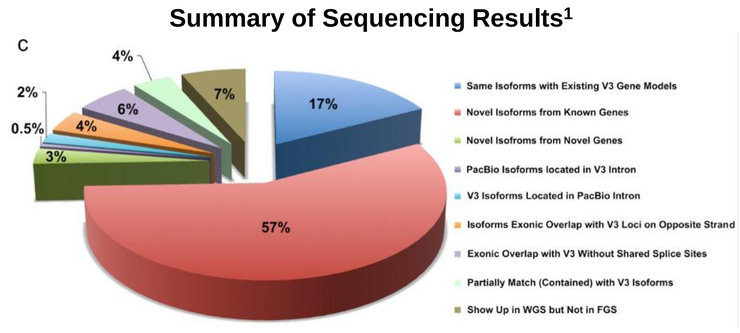
Distribution of full-length transcript sizes across all six tissues combined (left) or each tissue individually (right) using barcodes to segregate the samples. Libraries were non-size selected and run on the Sequel instrument.

Barcode Distribution

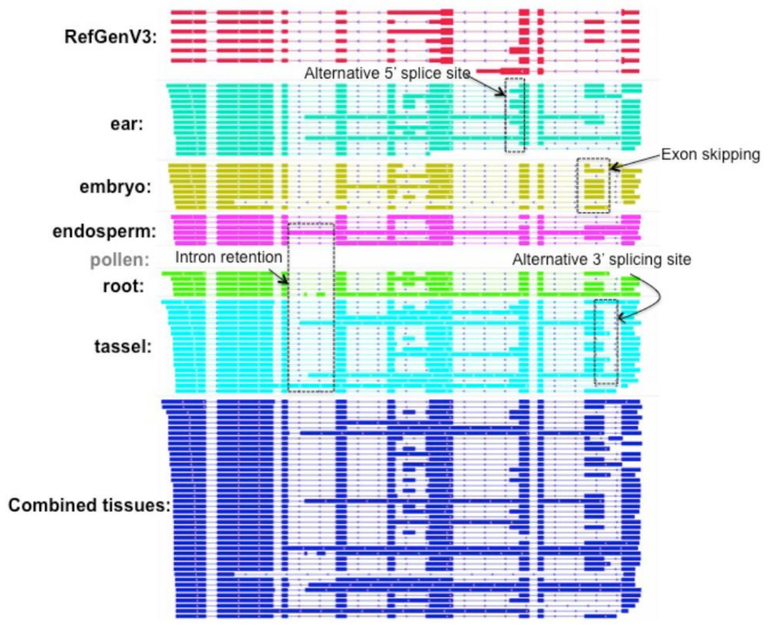


Count of each identified barcode for a single SMRT Cell run on either the PacBio RS II or Sequel instrument.

Summary of Sequencing Results[1]



IGV Visualization of Different Splicing Modes in One Gene[1]



Summary and Resources

- The Iso-Seq method provides full-length cDNA sequences without the need for assembly
- Decreased loading bias on the Sequel System makes it possible to streamline the Iso-Seq sample preparation process
- Sequencing of full-length transcripts identifies novel isoforms and improves gene annotations

Resources:

- Iso-Seq application and protocols: http://www.pacb.com/applications/rna-sequencing/
- Iso-Seq data analysis: http://www.pacb.com/products-and-services/analytical-software/smrt-analysis/analysis-applications/rna-sequencing/

References:

1. Wang B. et al (2016) Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nature Communications*. 7, 11708.