PacBio Americas User Group Meeting Sample Prep Workshop Breakout Session:

Iso-Seq Method and Analysis

June.27.2017 / http://programs.pacificbiosciences.com/l/1652/2017-03-25/3sn5p2

AGENDA

- **Introduction**
  - Iso-Seq Analysis Overview and Applications
  - Iso-Seq Analysis Experimental Design considerations
  - Iso-Seq Analysis Sample Preparation Procedures and Workflow
    - PacBio RS II *versus* Sequel System
    - Multiplexed Iso-Seq
    - Targeted Iso-Seq
    - Bacterial Iso-Seq
  - Iso-Seq Analysis Technical Resources
- **PacBio Scientific Conference Poster Presentations**
- **Q&A and Open Discussion**

CURRENT STATE OF TRANSCRIPT ASSEMBLY

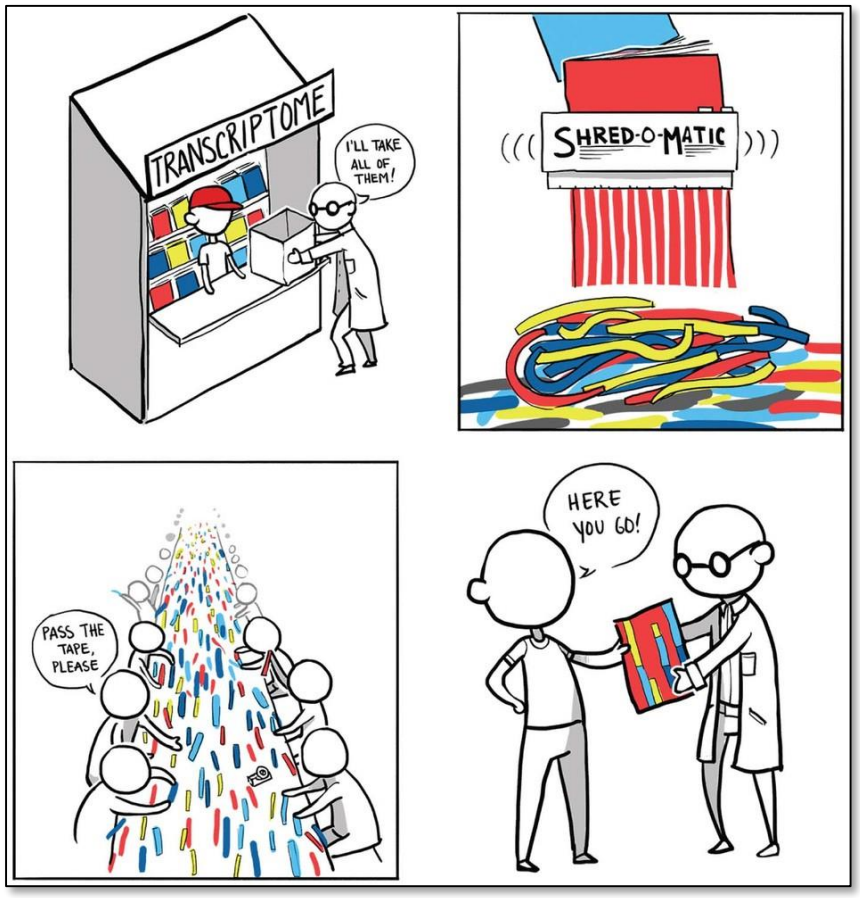


Figure 1 | Transcriptome reconstruction—akin to reassembling magazine articles after they have been through a paper shredder.

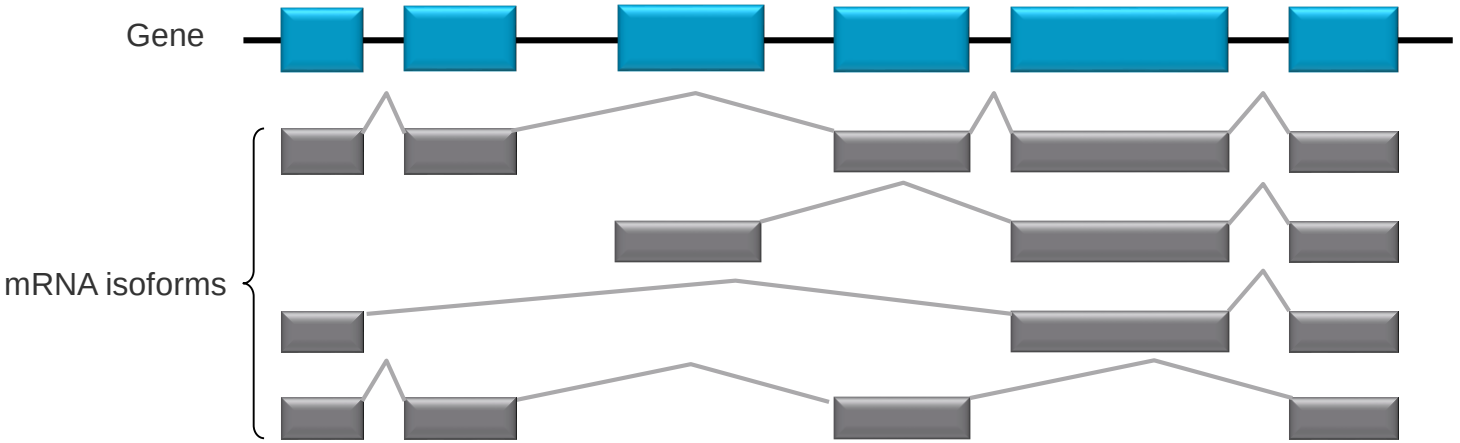
"The way we do RNA-seq now is…you take the transcriptome, you **blow it up into pieces** and then you try to figure out **how they all go back together again**…If you think about it, it's kind of a **crazy way to do things.**"

Michael Snyder
Stanford University

Tal Nawy (2013) End-to-end RNA sequencing, *Nature Methods* 10: 1144–1145

Ian Korf (2013) Genomics: the state of the art in RNA-seq analysis. *Nature Methods*. 10, 1165-1166

DETERMINATION OF TRANSCRIPT ISOFORMS

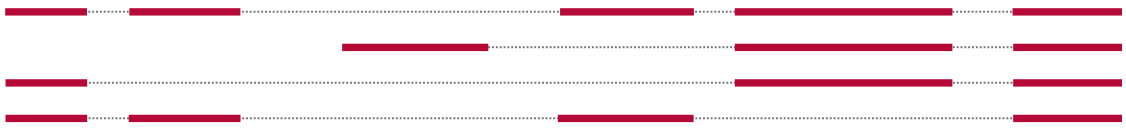


Short-read technologies:

Insufficient Connectivity
Splice Isoform Uncertainty

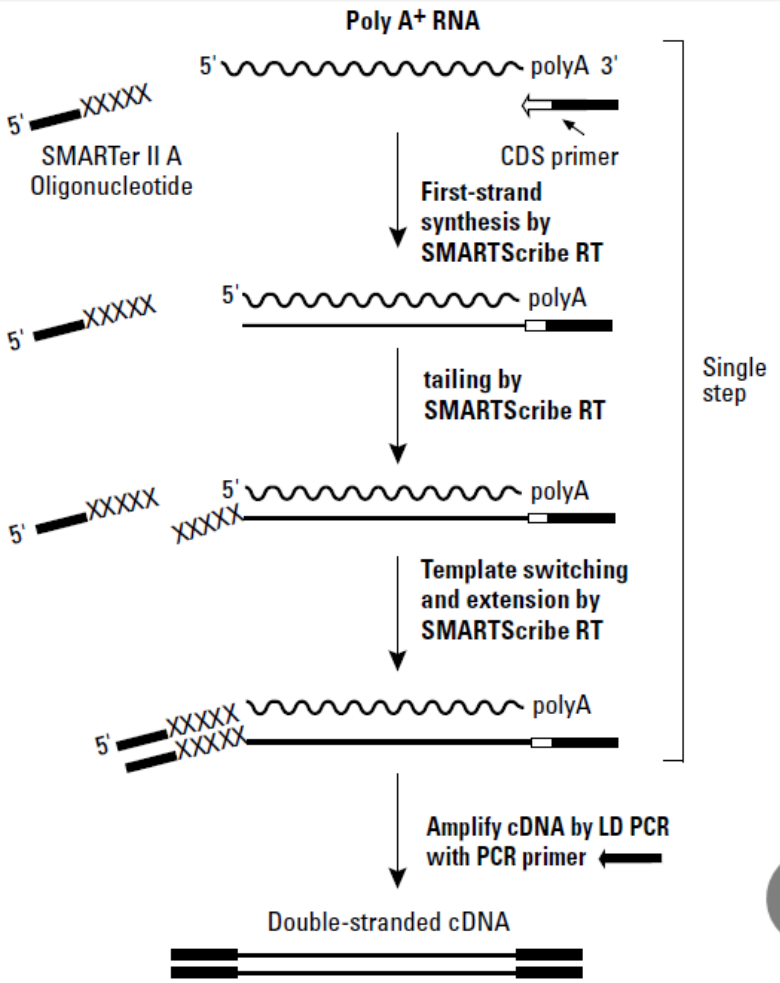
Reads spanning splice junctions

PacBio's Iso-Seq Analysis solution:



Full-length cDNA Sequence Reads
Splice Isoform Certainty – No Assembly Required

CLONTECH KIT: SELECTS FOR FULL LENGTH TRANSCRIPTS



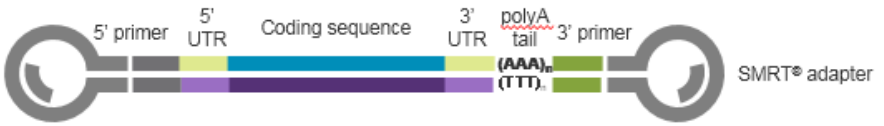
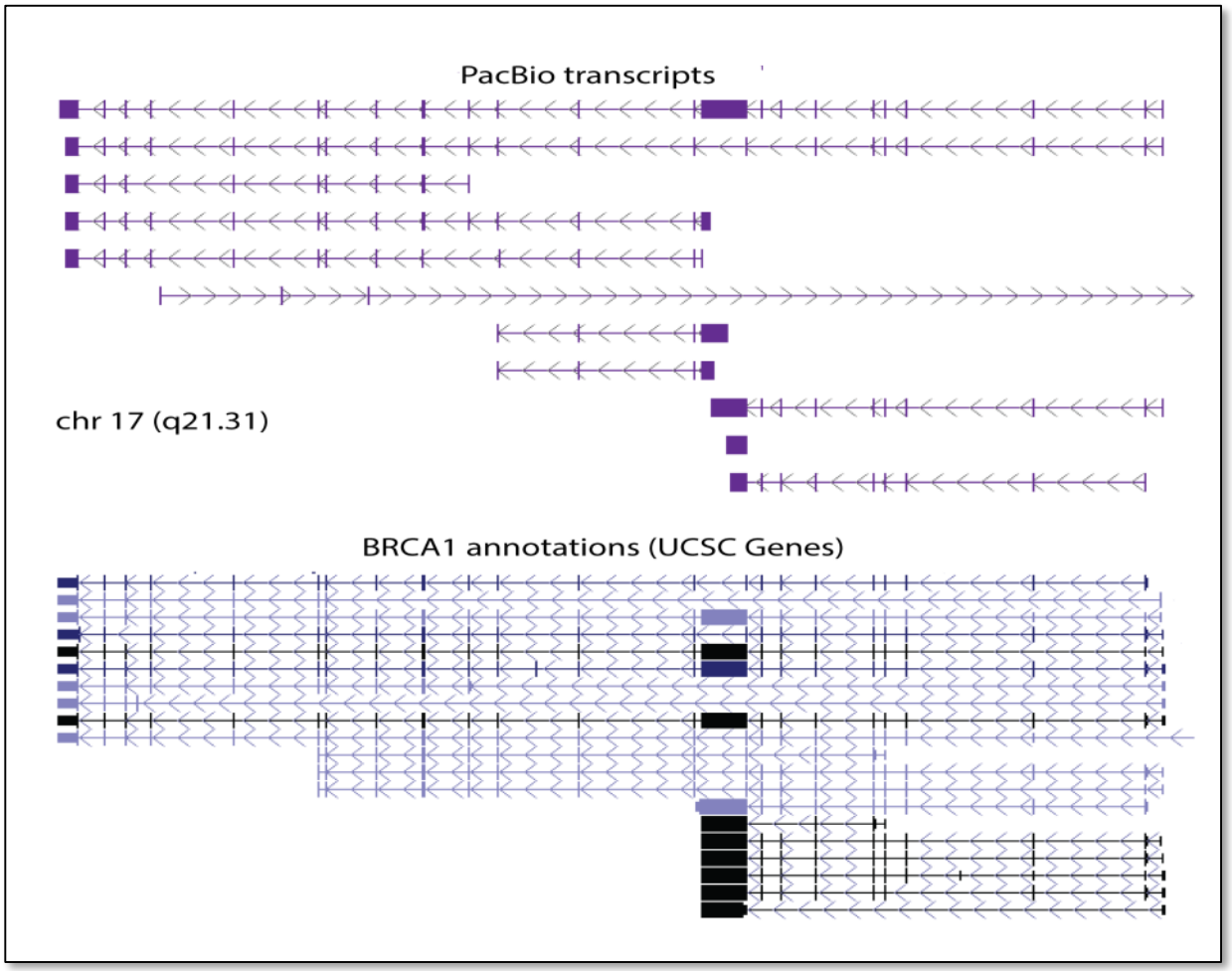


Figure 1. Flowchart of SMARTer cDNA synthesis. The SMARTer II A Oligonucleotide, 3′ SMART CDS Primer II A, and 5′ PCR Primer II A all contain a stretch of identical sequence (see Section I for sequence information).

BRCA1 ISOFORMS IN THE MCF-7 DATA

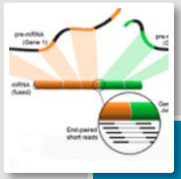


PacBio transcripts capture multiple isoforms of the BRCA1 gene, several of which are novel

ISO-SEQ ANALYSIS APPLICATIONS

Whole Transcriptome

- Genome Annotation
- Alternative Splice (De Novo & Reference)
- Full Length Isoform Discovery
- Biomarker Discovery
- RNA Capture Long Seq (CLS)
- Gene Expression



Targeted Iso-Seq

- Isoform Discovery
- Oncogenes
- Fusion Transcripts
- LNC RNA
- SV

Single Cell Iso-Seq

- Cell Differentiation
- Tissue Profiling

Iso-Seq Analysis Experimental Design Considerations

EXPERIMENTAL DESIGN CONSIDERATIONS

What are the goals of your application?

- Targeted or Full Transcriptome
- Alternative Splicing Analysis
- Gene Annotation

Is size selection needed? What size bins are required?

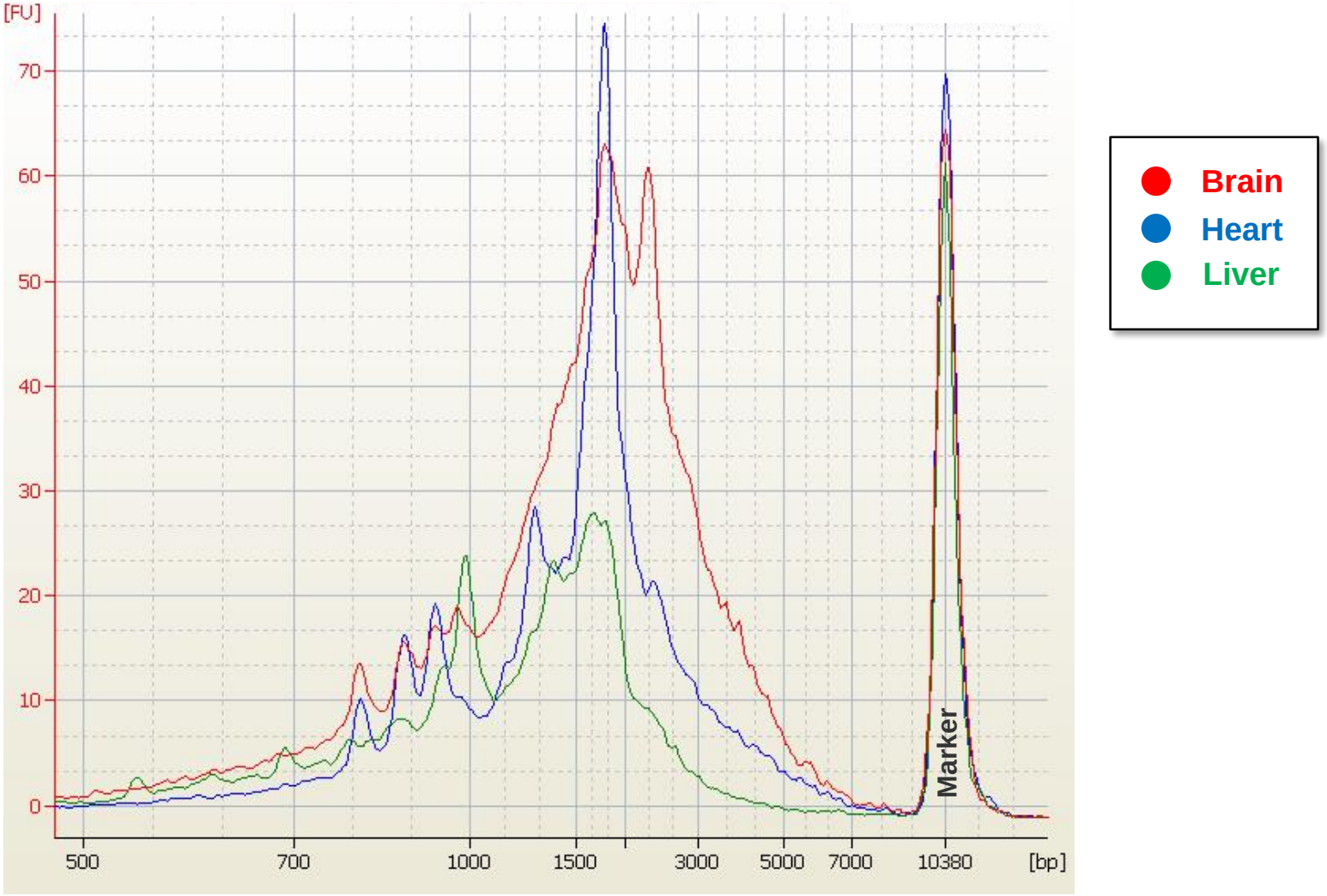
- Size selection: Yes/No
- Size selection method: Agarose Gel or Sage BluePippin or SageELF System

What are the estimated number of full-length transcripts? Is this enough to answer my scientific question?

- **PacBio RS II**: ~20,000 to 25,000 full-length transcript sequences per SMRT Cell
- **Sequel System:** ~200,000 to 250,000 full-length transcript sequences per SMRT Cell 1M
- Larger size fractions will have a lower percentage of FL reads

Required number of SMRT Cells is project and research-goal dependent, and depends strongly on transcriptome complexity of the organism being studied

AMPLIFIED CDNA LIBRARIES HAVE BROAD SIZE DISTRIBUTIONS



EXPERIMENTAL DESIGN RECOMMENDATIONS

Experimental Goals	# Sequel SMRT Cell	# RS II SMRT Cell
Targeted, gene-specific isoform characterization	<1	1
General survey of full-length isoforms in a transcriptome (moderate to high expression levels)	1 - 2	1 - 8
A comprehensive survey of full-length isoforms in the transcriptome	1 - 2	12 - 16
Deep sequencing for comprehensive isoform discovery and identification of low abundance transcripts	>3	>16

* For multiplexing, start with **1 Sequel SMRT Cell 1M per tissue** (e.g. if barcoding 6 tissues, start with ~6 Sequel SMRT Cells 1M)

Iso-Seq Analysis Sample Preparation Workflow: PacBio *RS II* versus *Sequel System*

RNA SAMPLE QUALITY REQUIREMENTS

For optimal sequencing performance, it is essential that the RNA sample:

- Has not undergone multiple freeze-thaw cycles as they can lead to RNA degradation.
- Has not been exposed to high temperatures or pH extremes (<6 or >9). For example, exposure to 70°C for >10 min or pH >8.5 will cause degradation.
- Has an OD260/OD280 ratio between 2.0 and 2.2.
- Has an OD260/OD230 ratio between 1.8 and 2.1.
- Has a RIN number ≥9 (Recommended).
- Does not contain insoluble material.
- Does not contain DNA contamination.
- Has not been exposed to intercalating fluorescent dyes or ultraviolet radiation. SYBR dyes are not RNA damaging, but do avoid ethidium bromide.
- Does not contain denaturants (e.g., guanidinium salts or phenol) or detergents (e.g., SDS or Triton-X100).
- Does not contain carryover contamination from the original organism/tissue (e.g., heme, humic acid, polyphenols, etc.).
- Store RNA is DEPC treated or nuclease free water at - 80°C.
- Best to freshly prepare total RNA before starting library preparation.
- Note: RNA samples should only be shipped on dry ice.

Please take all precautions against additional damages and ensure that the cDNA sample:

- Has not been exposed to the same types of damaging agents, conditions, and contaminants listed above for RNA input samples
- Has an OD260/OD280 ratio of 1.8 to 2.0.

ISOFORM SEQUENCING PROCEDURE & CHECKLISTS

PacBio RS II

Sequel System

No Size Selection

PACIFIC BIOSCIENCES®

Procedure & Checklist - Isoform Sequencing (Iso-Seq™) using the Clontech® SMARTer® PCR cDNA Synthesis Kit and No Size Selection

http://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-Isoform-Sequencing-Iso-Seq-using-the-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-and-No-Size-Selection.pdf

PACBIO®

Procedure & Checklist - Iso-Seq™ Template Preparation for Sequel™ Systems

http://www.pacb.com/wp-content/uploads/Procedure-Checklist-Iso-Seq-Template-Preparation-Sequel-Systems.pdf

With Size Selection

PACIFIC BIOSCIENCES®

Procedure & Checklist - Isoform Sequencing (Iso-Seq™) using the Clontech® SMARTer® PCR cDNA Synthesis Kit and Manual Agarose-gel Size Selection

http://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-Isoform-Sequencing-Iso-Seq-using-the-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-and-Manual-Agarose-gel-Size-Selection.pdf

PACIFIC BIOSCIENCES®

Procedure and Checklist - Isoform Sequencing (Iso-Seq™) Using the Clontech® SMARTer® PCR cDNA Synthesis Kit and BluePippin™ Size-Selection System

http://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-Isoform-Sequencing-Iso-Seq-using-the-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-and-the-BluePippin-Size-Selection-System.pdf

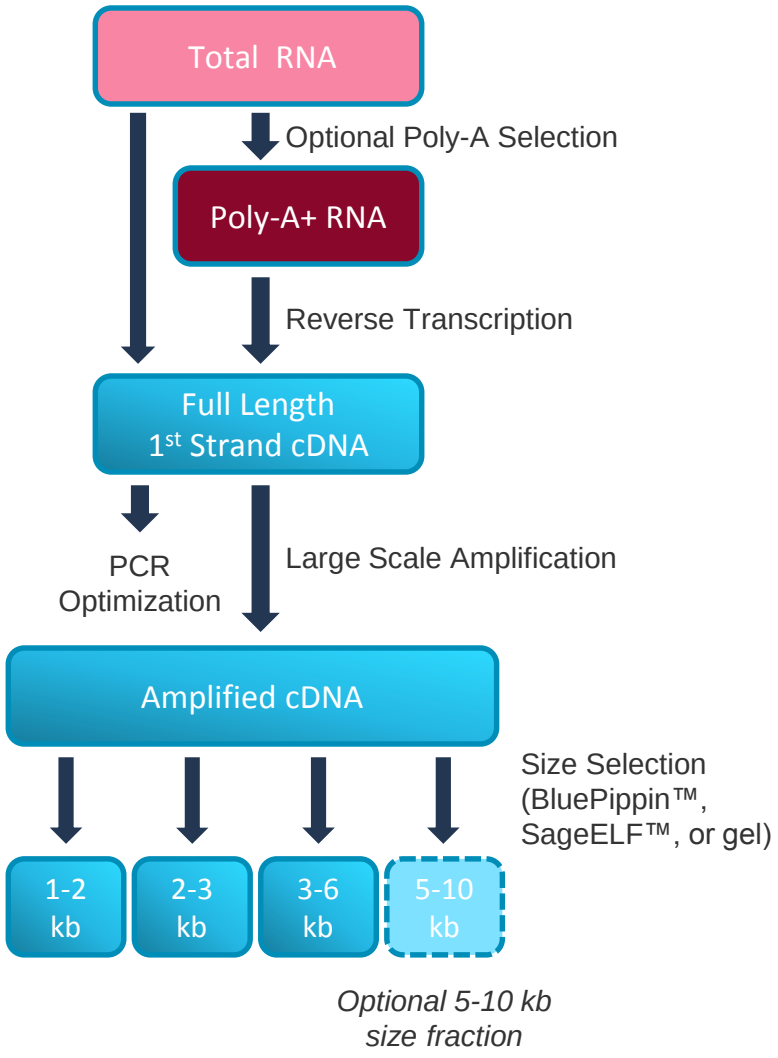
PACBIO®

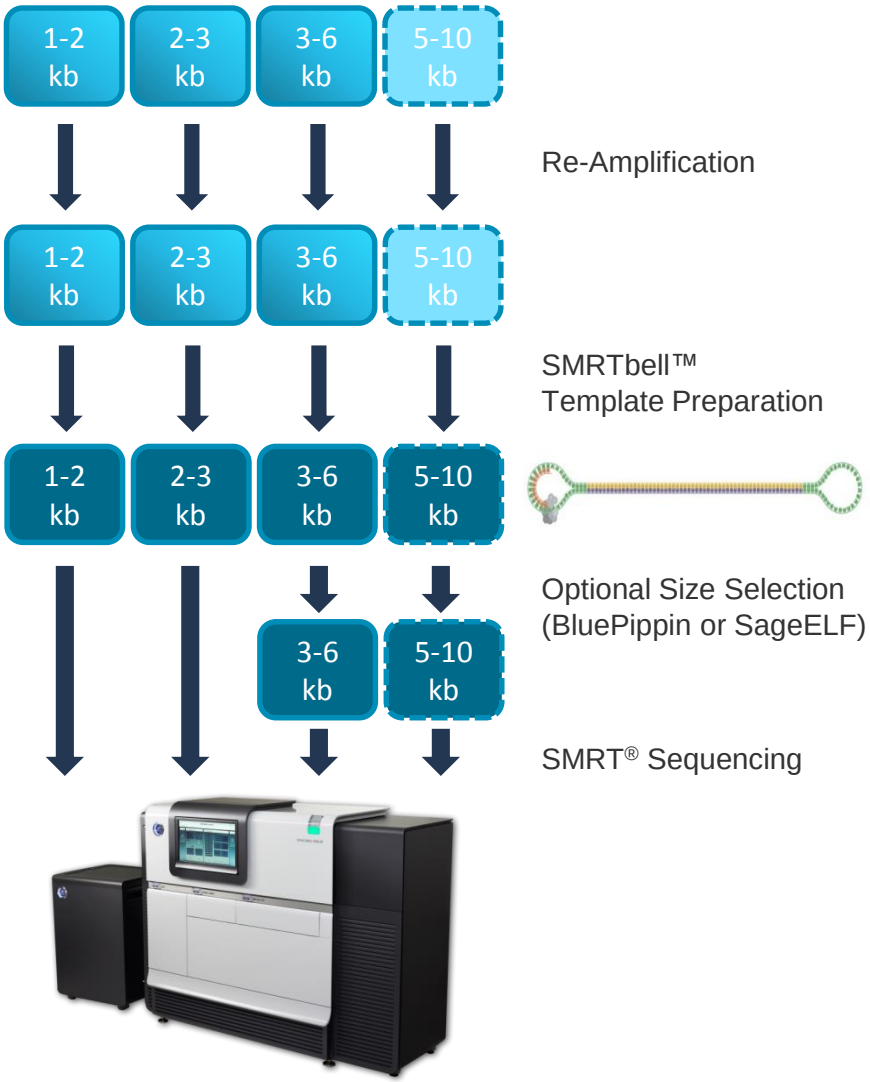
Procedure and Checklist - Isoform Sequencing (Iso-Seq™ Analysis) Using the Clontech SMARTer cDNA Synthesis Kit and SageELF™ Size-selection System

http://www.pacb.com/wp-content/uploads/Procedure-Checklist-Isoform-Sequencing-Iso-Seq-Analysis-using-the-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-and-SageELF-Size-Selection-System.pdf

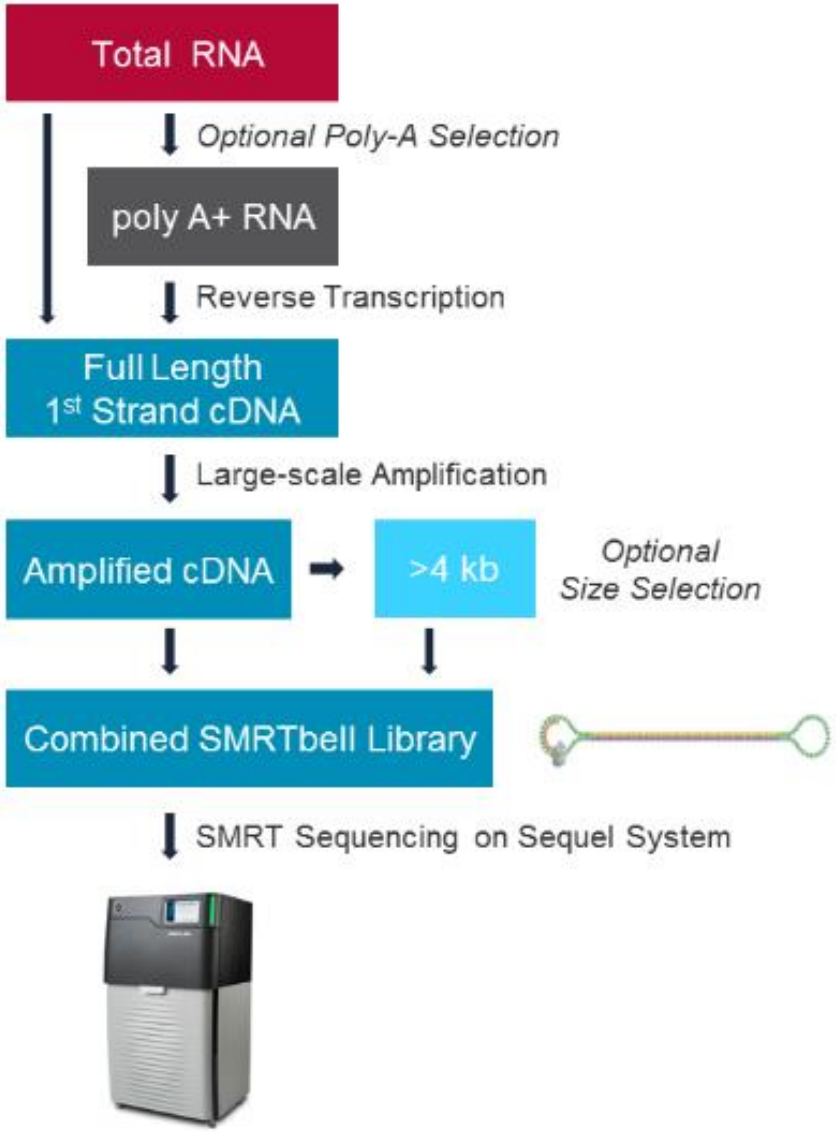
Find all protocols at http://www.pacb.com/support/documentation/

ISO-SEQ SAMPLE PREPARATION WORKFLOW OVERVIEW FOR PACBIO *RS II*



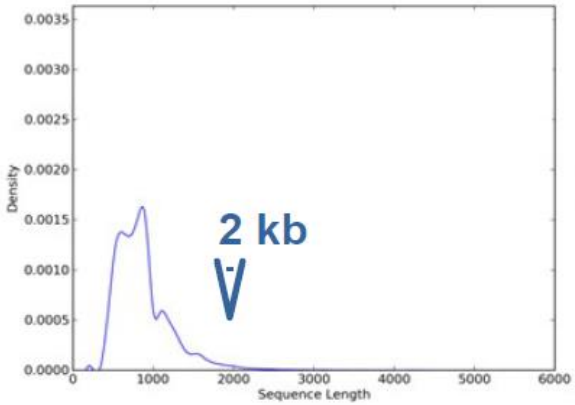


NEW STREAMLINED ISO-SEQ WORKFLOW FOR SEQUEL

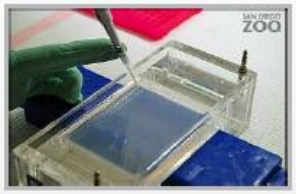


SIZE SELECTION ADVANTAGES ON THE PACBIO *RS II*

Transcript length distribution with no size-selection

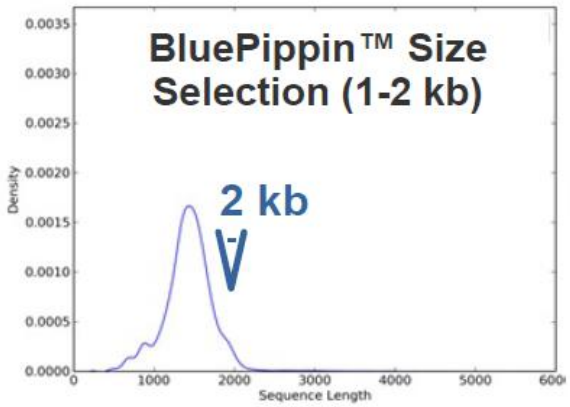


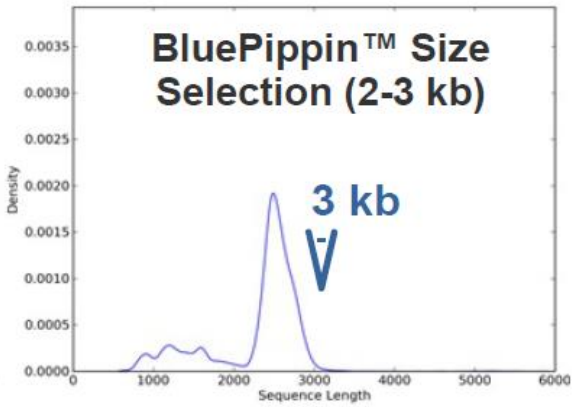
Size Selection Options:

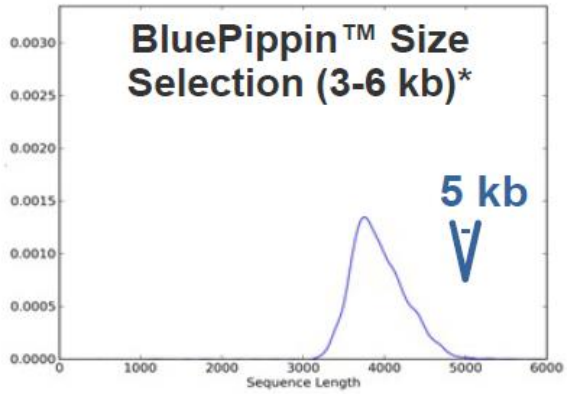


BluePippin™ System or manual agarose gel

Capture the broadest range of transcript lengths with size selection

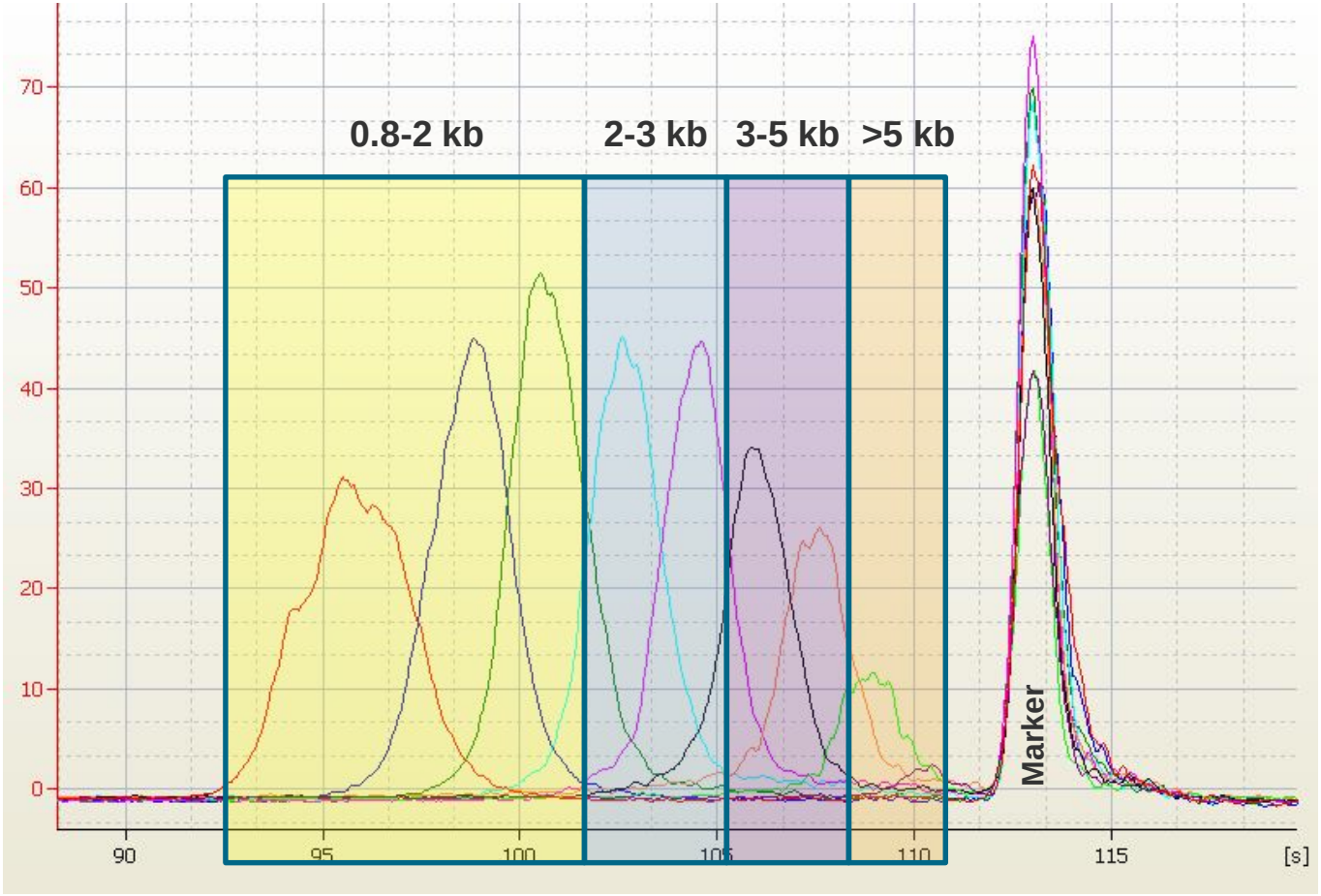






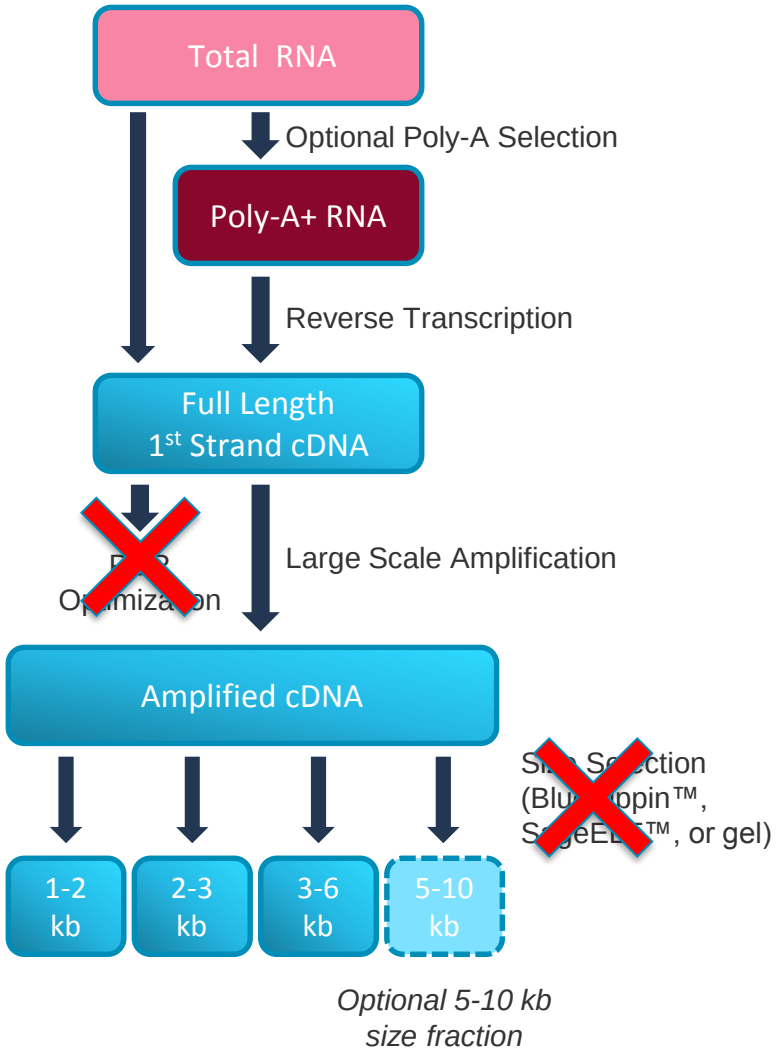
*3-6 kb bin uses two BluePippin size selections

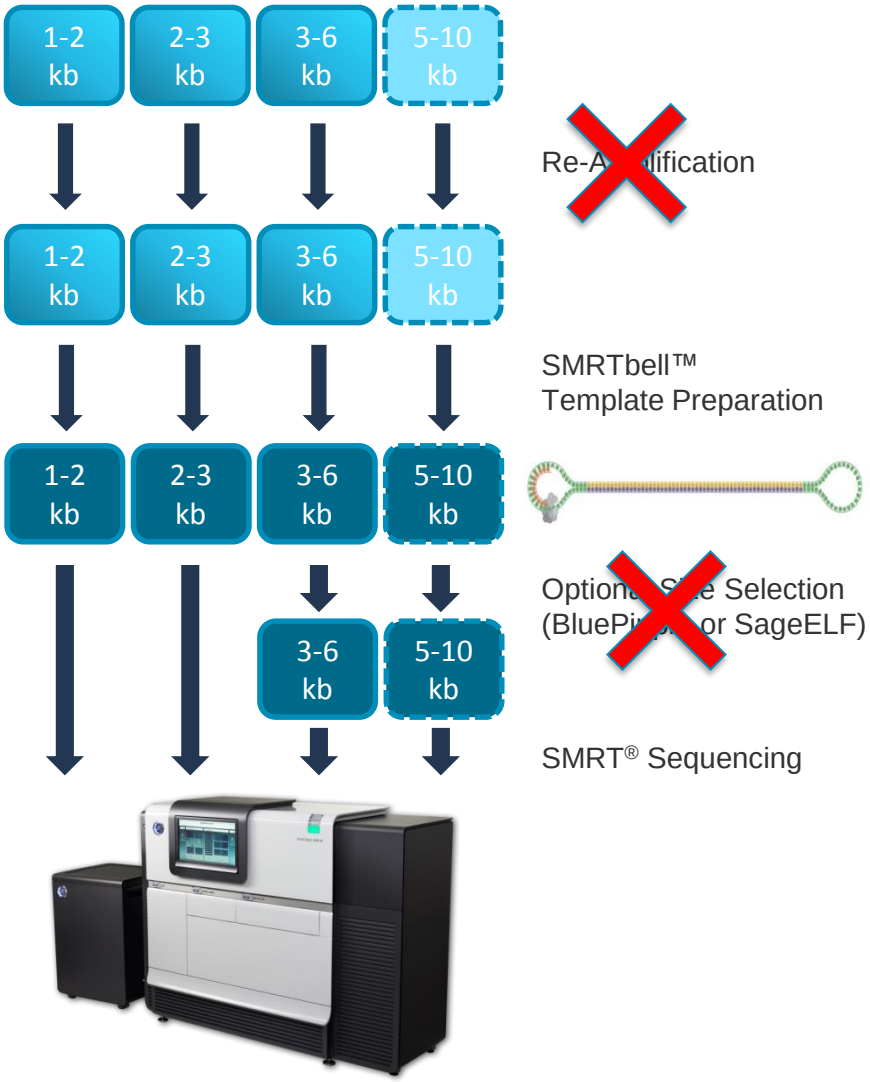
AMPLIFIED CDNA AFTER SIZE FRACTIONATION ON SAGEELF SYSTEM



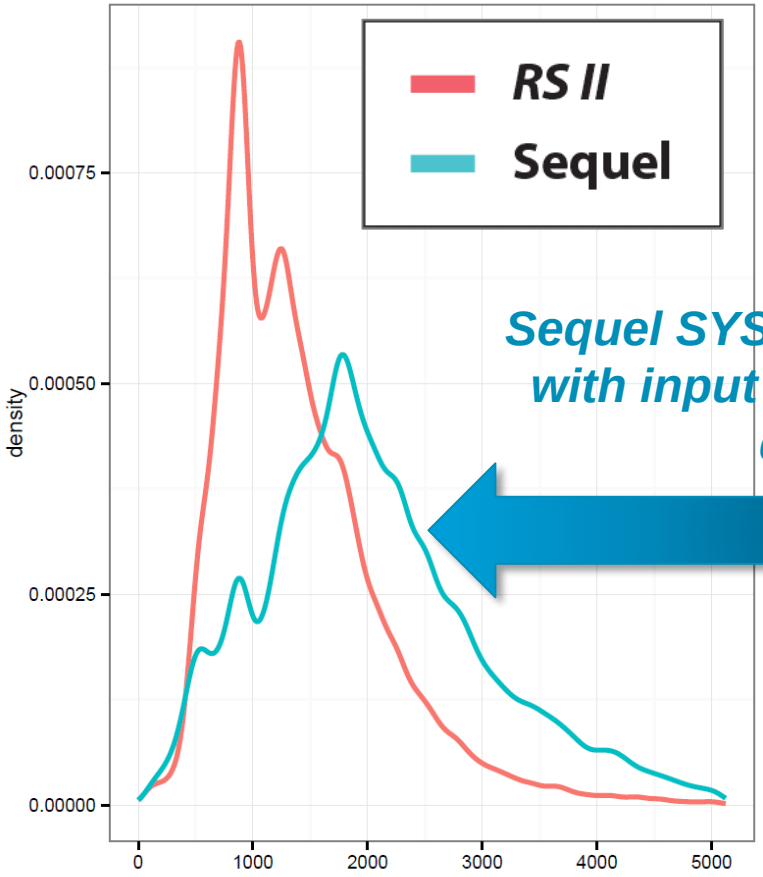
ELF	SIZE
11	1160
10	1650
9	1900
8	2400
7	2900
6	3700
5	5100
4	6050
3	7000

MODIFICATIONS TO ISO-SEQ SAMPLE PREPARATION WORKFLOW FOR *SEQUEL SYSTEM*



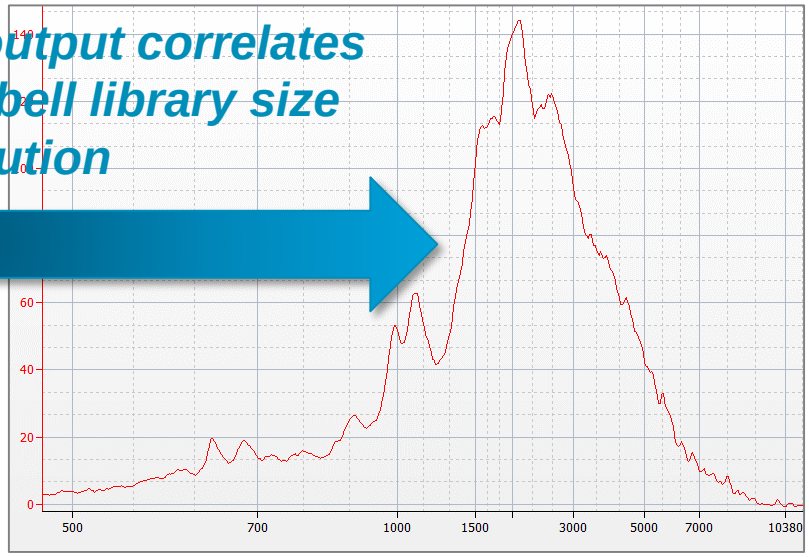


DECREASED LOADING BIAS IN SEQUEL SYSTEM REDUCES NEED FOR SIZE SELECTION



Full-length Transcript Sizes

Sequel SYSTEM output correlates with input SMRTbell library size distribution

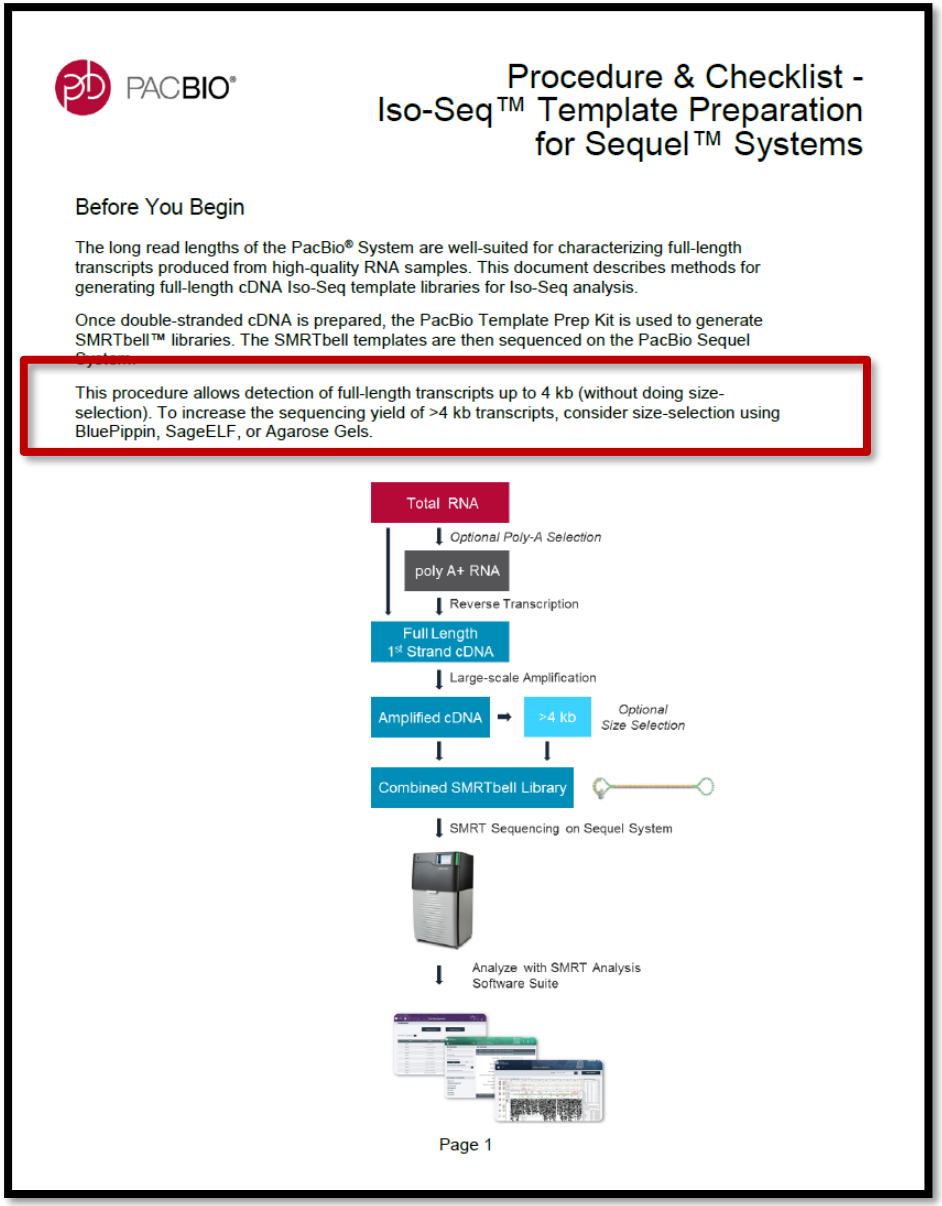


BioAnalyzer trace of a non-size selected Iso-Seq Library

- Histogram plot of number of full-length sequences by transcript length for a Magbead-loaded, non-size selected Iso-Seq library sequenced on both the PacBio RS II and the Sequel System.

- The full-length cDNA sequences run on the Sequel System closely resemble the size distribution of the input SMRTbell library

SIZE SELECTION IS OPTIONAL IN SEQUEL SYSTEM ISO-SEQ ANALYSIS



PACBIO®

Procedure & Checklist - Iso-Seq™ Template Preparation for Sequel™ Systems

Before You Begin

The long read lengths of the PacBio® System are well-suited for characterizing full-length transcripts produced from high-quality RNA samples. This document describes methods for generating full-length cDNA Iso-Seq template libraries for Iso-Seq analysis.

Once double-stranded cDNA is prepared, the PacBio Template Prep Kit is used to generate SMRTbell™ libraries. The SMRTbell templates are then sequenced on the PacBio Sequel System.

This procedure allows detection of full-length transcripts up to 4 kb (without doing size-selection). To increase the sequencing yield of >4 kb transcripts, consider size-selection using BluePippin, SageELF, or Agarose Gels.

Page 1

Size selection options to increase yield of >4 kb full-length transcript reads:

BluePippin:

- Select using 1 size fraction (4.5 – 10 kb)

SageELF:

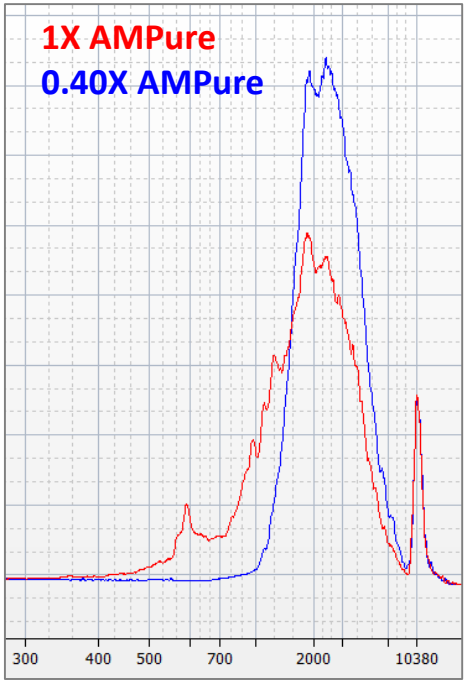
- Use existing protocol to size fractionate, but pool ONLY fractions #1 – 4

Manual Agarose Gel:

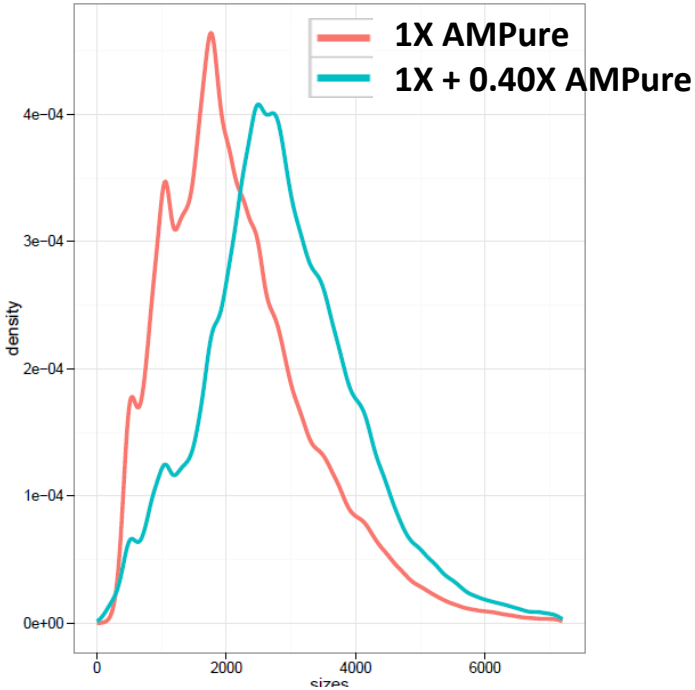
- Manually gel extract bands using a lower cutoff size of 4.5 kb

KEY CHANGES BETWEEN PACBIO RS II AND SEQUEL SYSTEM ISO-SEQ METHOD WORKFLOWS

- No size selection
- Updated PCR cycling reaction conditions for large-scale amplification of cDNA library
  - Use PrimeSTAR GXL (instead of Kapa HiFi) for improved amplification of longer transcripts
- After large-scale PCR amplification of cDNA library, split the sample and perform two parallel AMPure purifications using 0.40X and 1X concentrations
  - Pool the two purified aliquots in equimolar amounts after purification

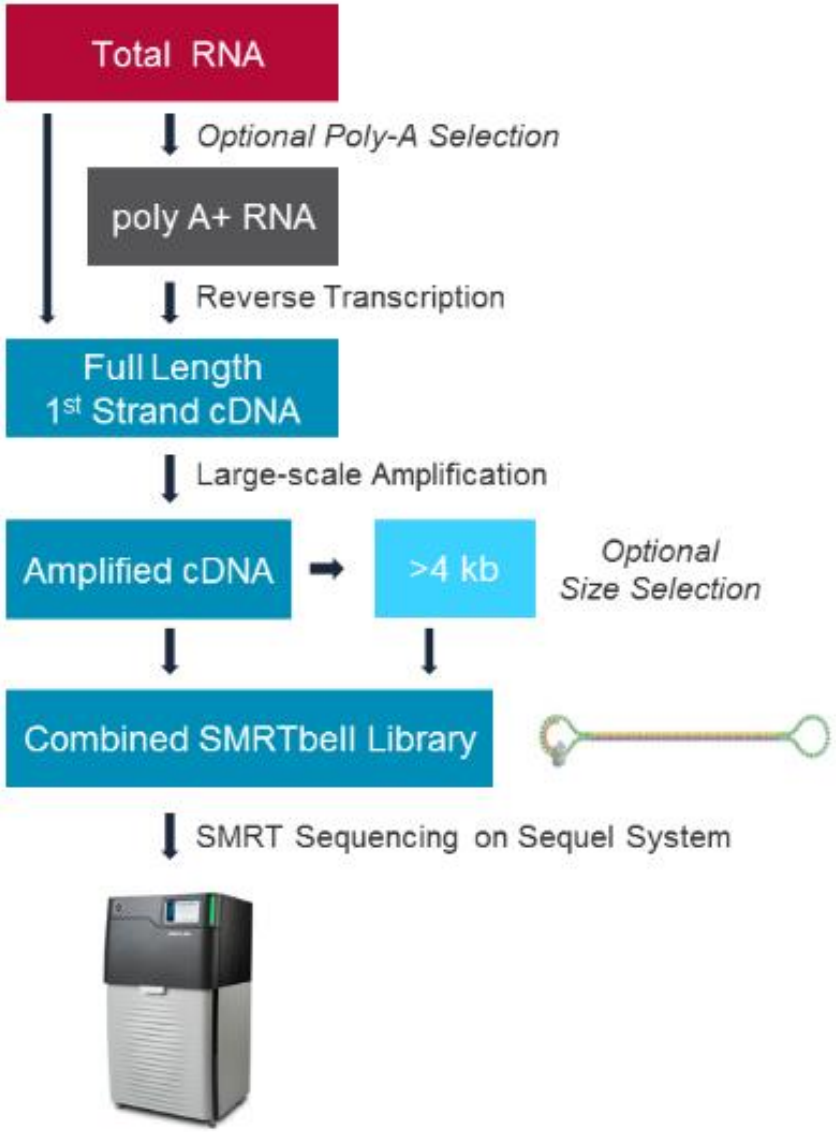


BioAnalyzer of 1X and 0.40X AMPure

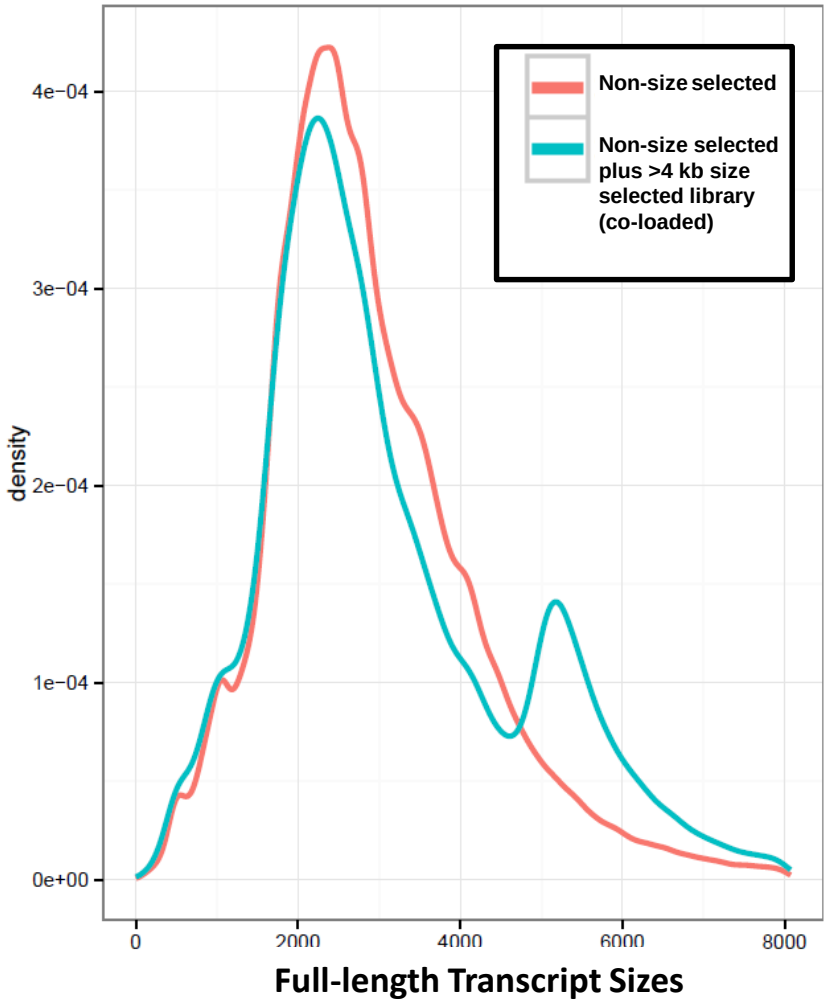


Full-length Transcript Size (bp)

NEW STREAMLINED ISO-SEQ WORKFLOW FOR SEQUEL



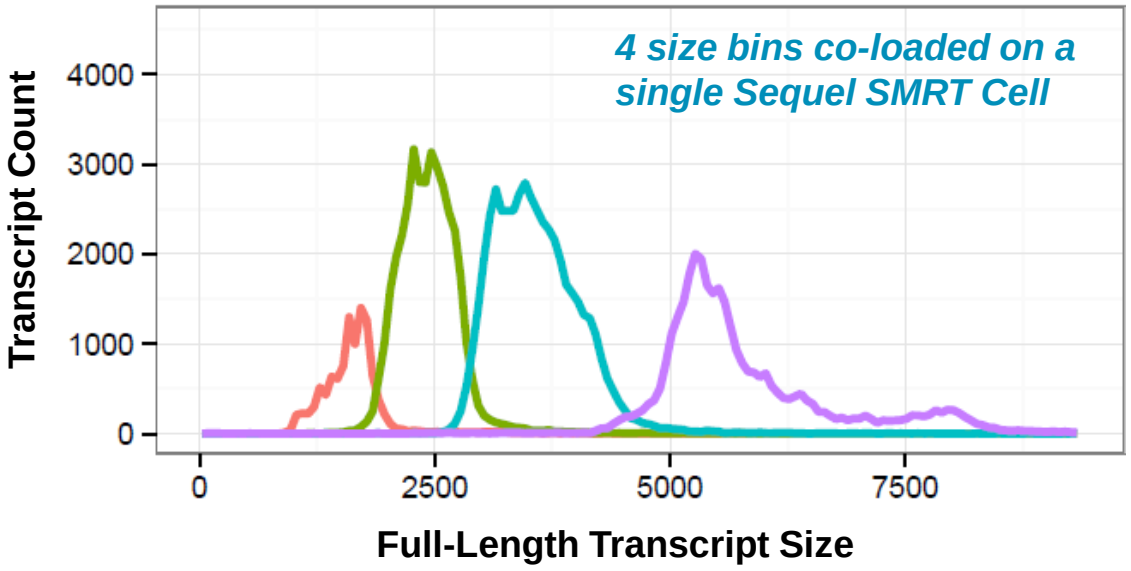
EFFECT OF USING SIZE SELECTION OPTION WITH SEQUEL SYSTEM ISO-SEQ ANALYSIS



- BluePippin (or SageELF) size-selected library (4.5 – 10 kb) can be pooled with non-size selected library and co-loaded together onto a single Sequel SMRT Cell

SUMMARY OF ISO-SEQ ANALYSIS WORKFLOW IMPROVEMENTS FOR SEQUEL SYSTEM

- Decreased requirement for size-selection
- Transcript sizes from non-size selected library represent input cDNA size distribution
- Multiple size bins can be co-loaded on a single Sequel SMRT Cell 1M
- Extra size-selection of SMRTbell libraries no longer required



SEQUENCING RECOMMENDATIONS FOR ISO-SEQ ANALYSIS ON SEQUEL SYSTEM

- Use MagBead Loading with default immobilization parameter settings
- On-plate concentration: 40 – 50 pM
- Aim for ≥ ~50% P1 yield (In comparison, PacBio RS II recommendation is P1 ≤ ~70%)
- Movie Time: 6 hours (10 hours may be helpful to further extend Polymerase Read Lengths)
  - Balance movie time (hours) vs instrument throughput (Gb)
- Enable Pre-extension Time = 120 min
- Spin column clean-up is not necessary
  - Effect of spin column is not yet fully understood
- If higher yield of >4 kb transcripts is desired, use size selection (4.5 – 10 kb)
- For Users who have existing Iso-Seq size-selected libraries, libraries may be pooled in equimolar ratio and loaded on a single Sequel SMRT Cell

Multiplexed Iso-Seq Analysis

MULTIPLEXED ISOFORM SEQUENCING PROCEDURE

PACIFIC BIOSCIENCES®

PacBio SampleNet – Shared Protocol

Please note: the shared protocols described herein may not have been validated by Pacific Biosciences and are provided as-is and without any warranty. Use of these protocols is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis using the PacBio® RS II system. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

Barcoding Samples for Isoform Sequencing (Iso-Seq™ Analysis)

http://www.pacb.com/wp-content/uploads/2015/09/Barcoding-Samples-for-Isoform-Sequencing-Iso-Seq-Analysis.pdf

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

nature COMMUNICATIONS

ARTICLE

Received 29 Oct 2015 | Accepted 20 Apr 2016 | Published 24 Jun 2016

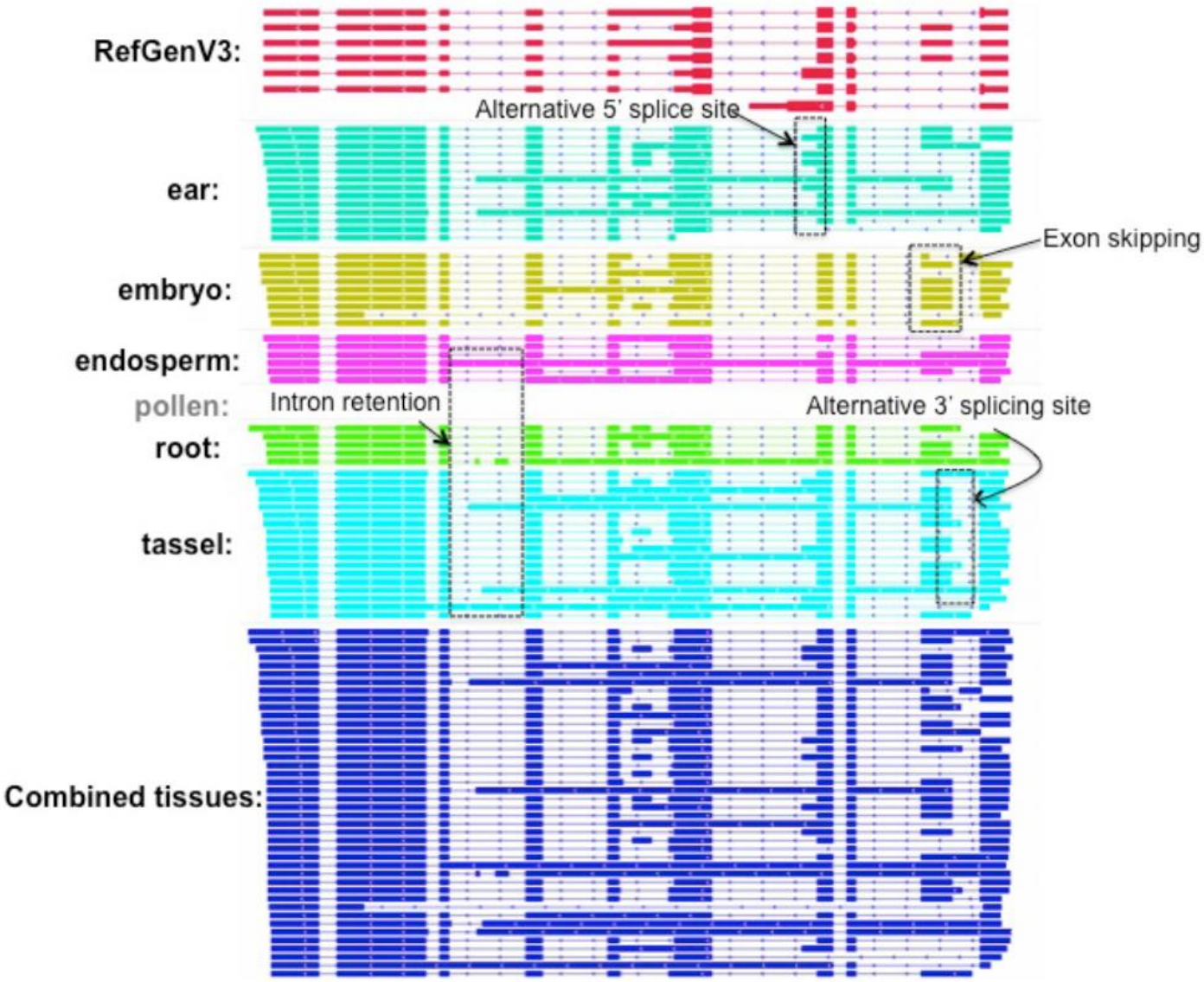
DOI: 10.1038/ncomms11708 OPEN

Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing

Bo Wang[1], Elizabeth Tseng[2], Michael Regulski[1], Tyson A. Clark[2], Ting Hon[2], Yinping Jiao[1], Zhenyuan Lu[1], Andrew Olson[1], Joshua C. Stein[1] & Doreen Ware[1,3]

BC	Tissue	1-2 kb	2-3 kb	3-5 kb	>5 kb
1	Root	17%	15%	14%	11%
2	Ear	13%	15%	17%	16%
3	Endosperm	16%	10%	9%	10%
4	Embryo	18%	21%	25%	34%
5	Tassel	15%	16%	17%	15%
6	Pollen	22%	23%	19%	15%

IGV Visualization of Different Splicing Modes in One Gene

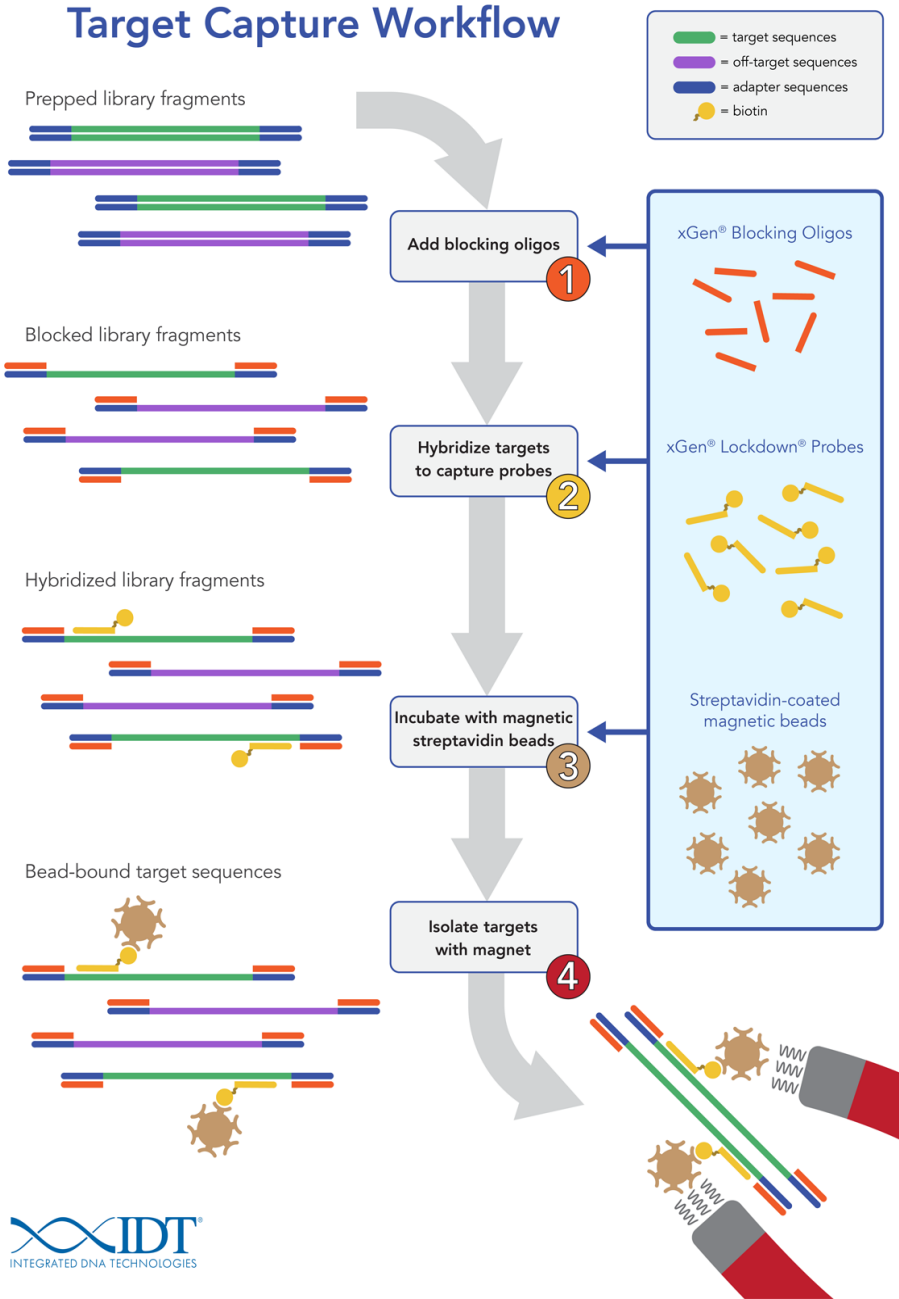


Targeted Iso-Seq Analysis

TARGETED FULL-LENGTH CDNA ISOFORM SEQUENCING USING PROBE-BASED CAPTURE

- RT-PCR

- Available kits for probe-based capture:
  - Nimblegen SeqCap
  - IDT
  - Agilent SureSelect



FULL-LENGTH CDNA TARGET SEQUENCE CAPTURE PROCEDURES

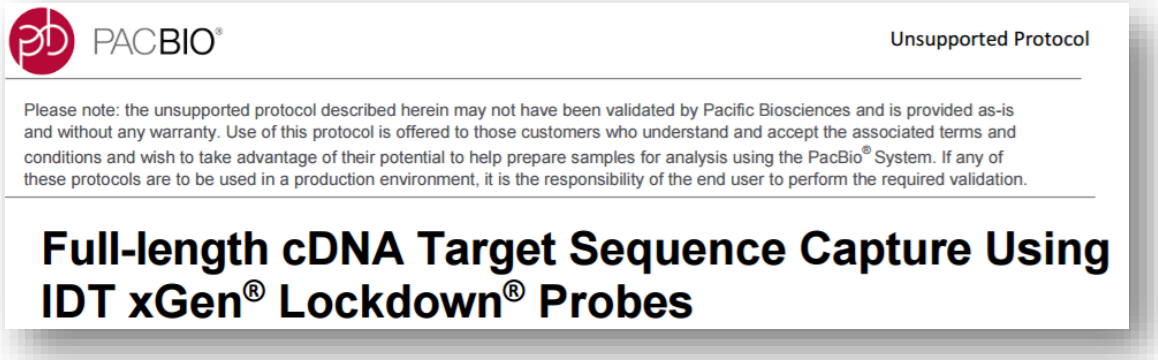

PACIFIC BIOSCIENCES®

PacBio SampleNet – Shared Protocol

Please note: the shared protocols described herein may not have been validated by Pacific Biosciences and are provided as-is and without any warranty. Use of these protocols is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis using the PacBio® system. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

Full-length cDNA Target Sequence Capture Using SeqCap® EZ Libraries

http://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Full-length-cDNA-Target-Sequence-Capture-Using-Roche-NimbleGen-SeqCap-EZ-Library.pdf


PACBIO®

Unsupported Protocol

Please note: the unsupported protocol described herein may not have been validated by Pacific Biosciences and is provided as-is and without any warranty. Use of this protocol is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis using the PacBio® System. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

Full-length cDNA Target Sequence Capture Using IDT xGen® Lockdown® Probes

http://www.pacb.com/wp-content/uploads/Unsupported-Protocol-Full-length-cDNA-Target-Sequence-Capture-IDT-xGen-Lockdown-Probes.pdf

Multiplexed targeted Iso-Seq analysis can be achieved by incorporating barcodes early in the workflow:

- Reduces total time to sample Prep
- Reduces amplification, size selection and SMRTbell template preparation cost

Find all protocols at http://www.pacb.com/support/documentation/

A Method for the Identification of Variants in Alzheimer's Disease Candidate Genes and Transcripts Using Hybridization Capture Combined with Long-Read Sequencing

Steve Kujawa[1], Jenny Ekholm[1], Kevin Eng[1], Ting Hon[1], Elizabeth Tseng[1], Aaron Wenger[1], Kristina Giorda[2], Jiashi Wang[2] & Mirna Jarosz[2]
[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025
[2]Integrated DNA Technologies, 1710 Commercial Park, Coralville, IA, 52241

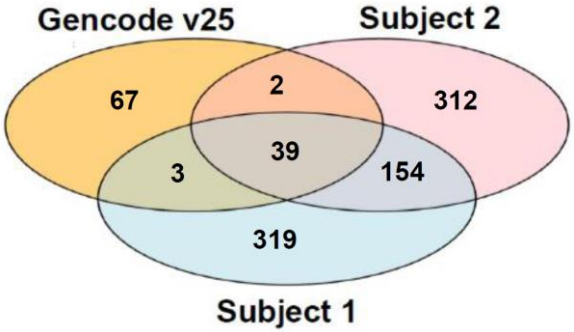
http://www.pacb.com/wp-content/uploads/Kujawa-AGBT-2017-Alzheimers-Disease-Candidate-Genes-and-Transcripts-Using-Hybridization-Capture.pdf

AGBT 2017

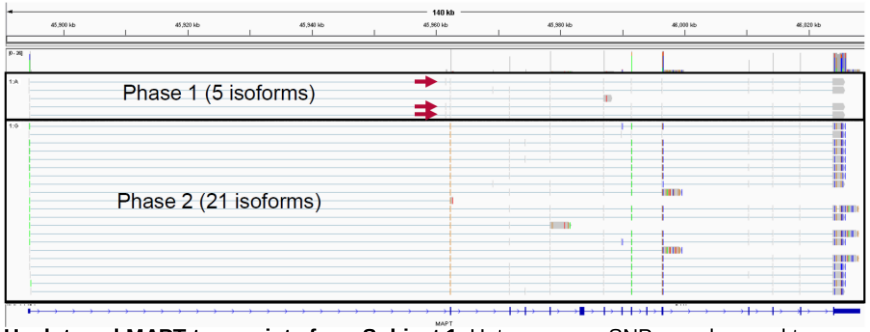
- Here we present a method for capturing genomic DNA (gDNA) and cDNA from two AD subjects using a panel of probes targeting 35 AD candidate genes
  - Probes were placed approximately every 1 kb and designed to cover the entire gene (exons, introns and regulatory regions)
- By combining xGen® Lockdown® probes with SMRT Sequencing, we provide completely sequenced candidate genes as well as their corresponding full-length transcripts.
- Furthermore, we are able to take advantage of heterozygous variants to phase the genes and their corresponding transcript isoforms into their respective haplotypes
- Combining xGen Lockdown probes with SMRT Sequencing provides a method for completely sequenced candidate genes and their corresponding full-length transcripts. This method enables:
  - Detection of a broad range of genomic variants, from SNPs to multi-kilobase insertions and deletions
  - Detection of novel transcript isoforms, including novel exons
  - Assignment of variants and transcripts isoforms to their specific haplotypes

	Subject	Source of Genomic DNA	Source of Total RNA
#1	87 year-old male	Brain, Frontal Lobe	Brain, Temporal Lobe
#2	93 year-old female	Skeletal Muscle	Brain, Temporal Lobe

gDNA and total RNA from two AD subjects were purchased from BioChain Institute, Inc.



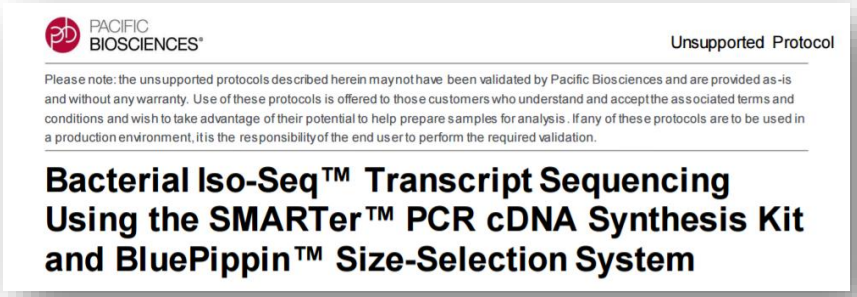
Comparison of isoforms observed in Subjects 1 & 2 with Level 1 isoforms in Gencode v25. In total, 515 and 507 isoforms were found in Subject 1 and 2, respectively. When compared with existing transcripts in Gencode v25 only 39 were shared among all 3 data sources



Haplotyped MAPT transcripts from Subject 1. Heterozygous SNPs can be used to haplotype the transcripts. A novel exon (red arrows) was observed in three of the five isoforms in Phase 1 and not observed in any of the 21 isoforms in Phase 2.

Bacterial Iso-Seq Analysis

BACTERIAL ISO-SEQ METHOD



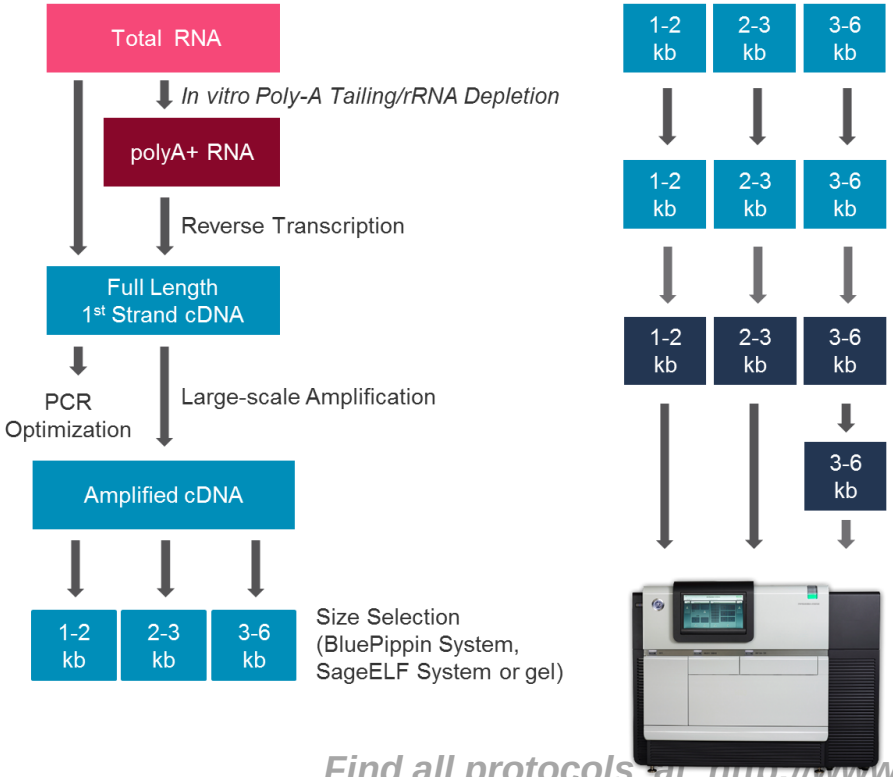
PACIFIC BIOSCIENCES®

Unsupported Protocol

Please note: the unsupported protocols described herein may not have been validated by Pacific Biosciences and are provided as-is and without any warranty. Use of these protocols is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

Bacterial Iso-Seq™ Transcript Sequencing Using the SMARTer™ PCR cDNA Synthesis Kit and BluePippin™ Size-Selection System

http://www.pacb.com/wp-content/uploads/Unsupported-Protocol-Bacterial-Iso_Seq-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-BluePippin-Size-Selection.pdf

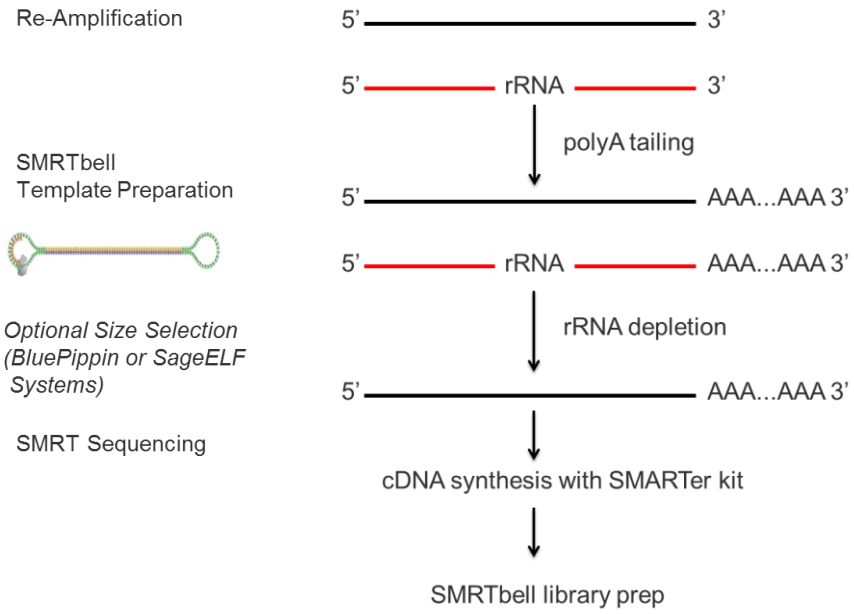


Goal

- Detection of polycistronic and full-length operon transcripts
- Detection of alternative transcription start/stop sites (TSS)

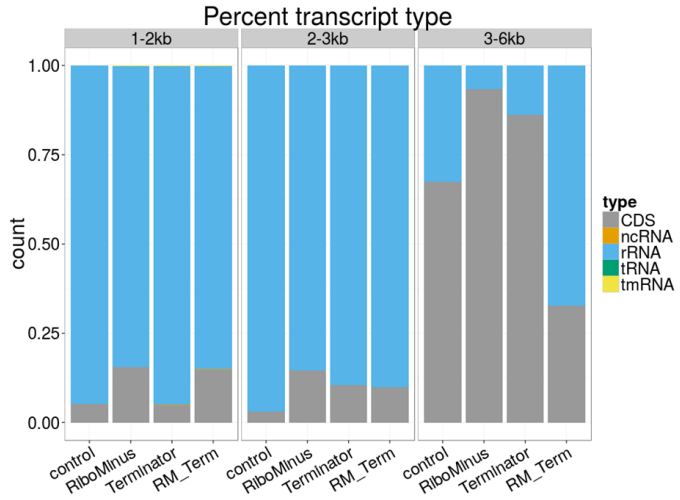
Bacterial Iso-Seq Analysis Considerations

- No polyA tail
- No built-in selection of mRNA
- To address the above challenges:
  - Enzymatically add poly(A)-tail
  - Perform rRNA depletion (with RiboMinus)

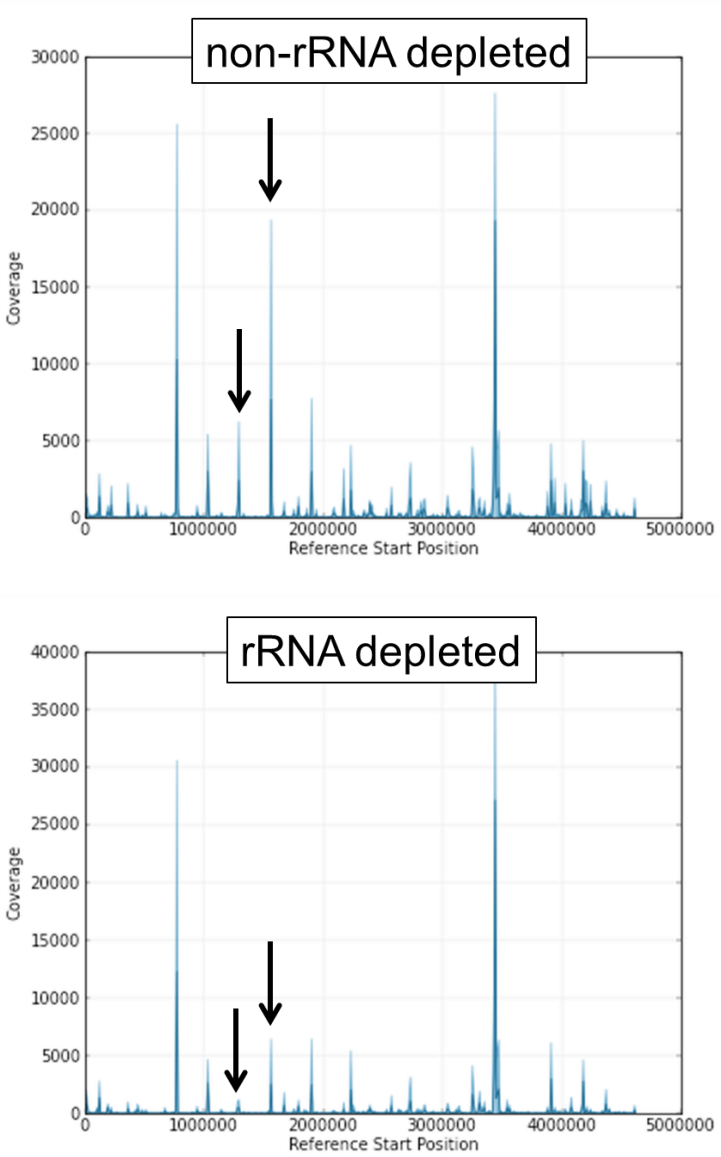


PREPARING BACTERIAL RNA FOR REVERSE TRANSCRIPTION: POLY(A)-TAILING AND RIBOSOMAL RNA DEPLETION

- The reverse transcription kit utilized for this method, the SMARTer PCR cDNA Synthesis Kit from Clontech, requires the presence of a poly(A)-tail in order to generate full-length cDNA
- Due to the absence of a poly(A)-tail on bacterial transcripts, users must first enzymatically add poly(A)-tails to 3' ends of all transcripts (using the Poly(A) Polymerase Tailing Kit from Epicentre).
- Since this will add poly(A)-tails to all transcripts, including highly abundant rRNA, users must then deplete rRNA using an rRNA-depletion method (RiboMinus Technology from Thermo Fisher is described in this protocol).



Sequence reads were mapped to the *E. coli* genome reference. Arrows show reduction in coverage of rRNAs after rRNA depletion. Shared peaks are most likely ribosomal associated genes. Out of all traditional rRNA depletion methods tested, RiboMinus had the highest rRNA depletion efficiency.



Poster #687

Full-length cDNA Sequencing of Prokaryotic Transcriptome and Metatranscriptome Samples

Matthew Boitano[1], Bo Yan[2], Ting Hon[1], Elizabeth Tseng[1], Laurence Ettwiller[2] and Tyson A. Clark[1]
[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025
[2]New England Biolabs, 240 County Road, Ipswich, MA 01938

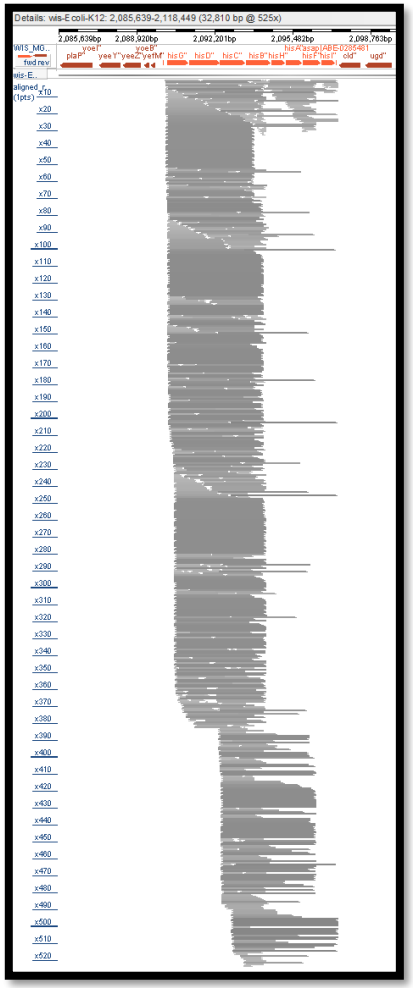
PacBio
(650) 521- 8342
mboitano@pacb.com

PDF Available Upon Request

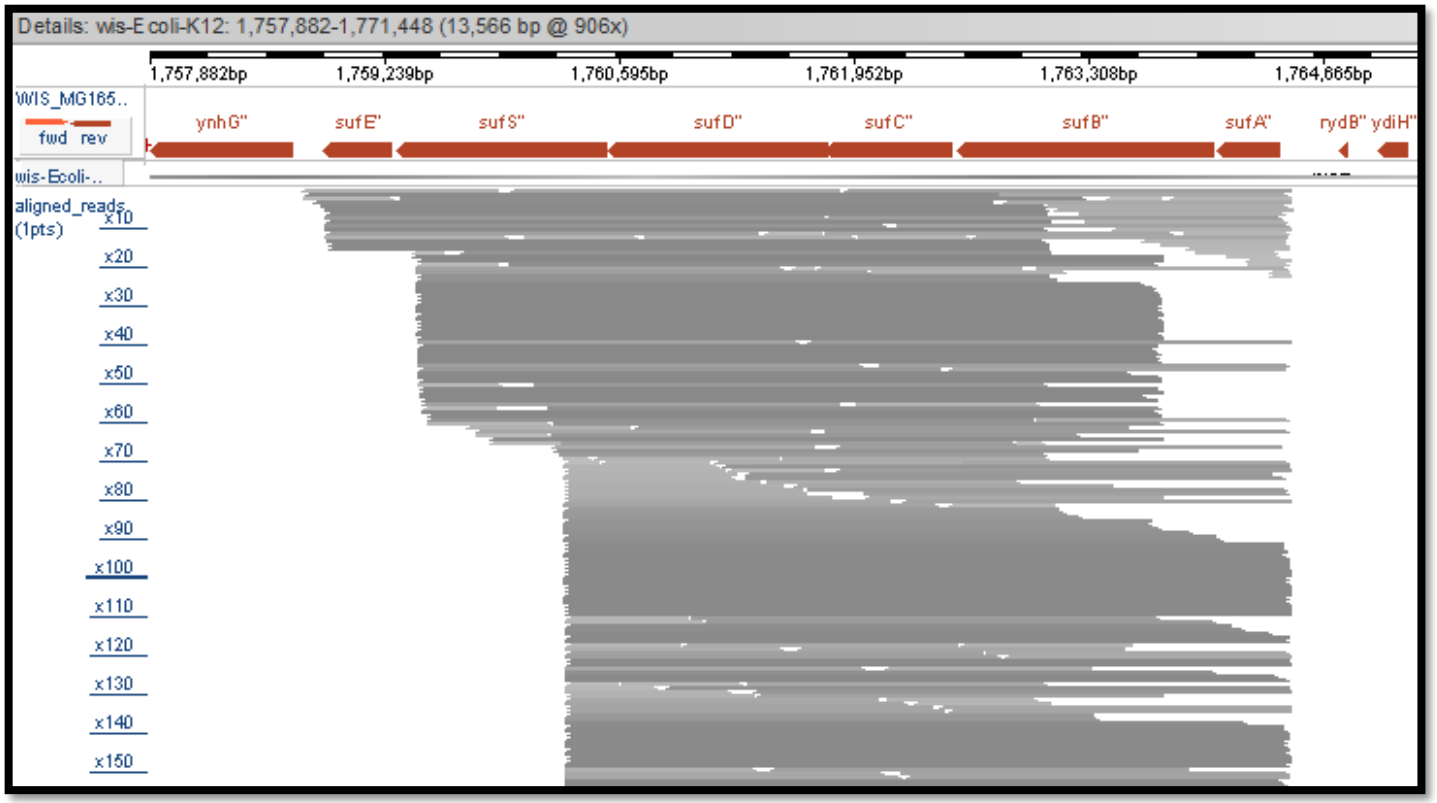
ASM 2017

Using long-read SMRT Sequencing, poly-cistronic and full-length operon reads are easily obtained without the need for assembly of short fragments

(A)



(B)



- Using long-read SMRT Sequencing, distinct transcription start/stop sites can be identified. Full-length transcript reads that map to the **(A)** his and **(B)** sur operons in *E. coli* show multiple transcription start and stop sites, resulting in multiple, distinct transcripts from the same operon.

PACBIO USER GROUP MEETING AGENDA – JUNE 28, 2017

11:50 – 12:20 p.m.

SMRT-cappable-seq Reveals the Complex Operome of Bacteria

Bo Yan, Ph.D. Postdoc, New England Biolabs (NEB)

Gene expression in bacteria is organized into operons, a functional unit of genomic DNA containing a cluster of genes under the control of a single promoter. Operon structures in bacteria are difficult to tackle because the accurate identification of transcripts start and end is challenging. Additionally, short read sequencing cannot establish the connectivity between the transcription start and end. As a result, and despite overwhelming mass of transcriptome information, accurate operon structures have so far been solved only for a handful of cases.

Here we describe SMRT-cappable-seq as the first experimental methodology to identify operon structures genome-wide in prokaryotes. It combines the isolation of full length primary transcripts with SMRT Sequencing. Applied to *E. coli*, SMRT-cappable-seq identifies a total of 2300 operons from which around 900 are novel. Importantly, our result reveals a pervasive read-through of previous experimentally validated transcription termination sites. Termination read-through represents a powerful strategy to control gene expression and establish operon polarity. Taken together this data provides a first glance at the complexity of the ‘operome’ in bacteria and presents an invaluable resource for understanding gene regulation and function in bacteria.

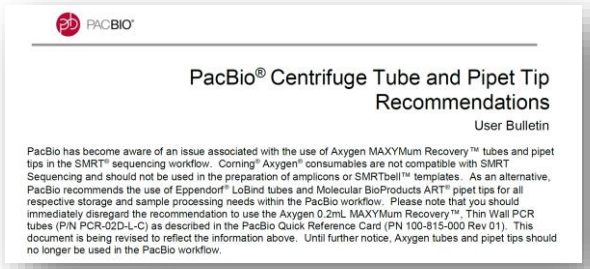
Iso-Seq Technical Resources

TECHNICAL RESOURCES FOR SMRTBELL LIBRARY SAMPLE PREPARATION AND ISO-SEQ ANALYSIS

User Bulletins

User Bulletin for PacBio RS II and Sequel Systems: Centrifuge Tube and Pipet Tip Recommendations ***(NEW!)*** *(May 2017)*

- PacBio advises against the use of Axygen MAXYMum Recovery™ tubes and pipet tips. Please discontinue use of these products immediately. PacBio recommends alternatives in the User Bulletin.
- http://www.pacb.com/wp-content/uploads/User-Bulletin-Centrifuge-Tube-and-Pipet-Tip-Recommendations.pdf

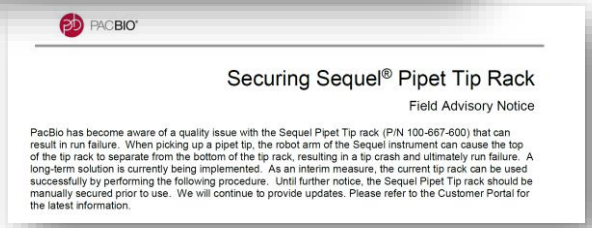

PacBio® Centrifuge Tube and Pipet Tip Recommendations

User Bulletin

PacBio has become aware of an issue associated with the use of Axygen MAXYMum Recovery™ tubes and pipet tips in the SMRT® sequencing workflow. Corning® Axygen® consumables are not compatible with SMRT Sequencing and should not be used in the preparation of amplicons or SMRTbell™ templates. As an alternative, PacBio recommends the use of Eppendorf® LoBind tubes and Molecular BioProducts ART® pipet tips for all respective storage and sample processing needs within the PacBio workflow. Please note that you should immediately disregard the recommendation to use the Axygen 0.2mL MAXYMum Recovery™, Thin Wall PCR tubes (P/N PCR-02D-L-C) as described in the PacBio Quick Reference Card (PN 100-815-000 Rev 01). This document is being revised to reflect the information above. Until further notice, Axygen tubes and pipet tips should no longer be used in the PacBio workflow.

Field Advisory for Sequel System: Securing Sequel Pipet Tip Rack ***(NEW!)*** *(May 2017)*

- PacBio recommends a simple procedure to ensure that the Sequel Pipet Tip rack is firmly affixed to the tip box.
- http://www.pacb.com/wp-content/uploads/Field-Advisory-Notice-Securing-Sequel-Pipet-Tip-Rack.pdf


Securing Sequel® Pipet Tip Rack

Field Advisory Notice

PacBio has become aware of a quality issue with the Sequel Pipet Tip rack (P/N 100-667-600) that can result in run failure. When picking up a pipet tip, the robot arm of the Sequel instrument can cause the top of the tip rack to separate from the bottom of the tip rack, resulting in a tip crash and ultimately run failure. A long-term solution is currently being implemented. As an interim measure, the current tip rack can be used successfully by performing the following procedure. Until further notice, the Sequel Pipet Tip rack should be manually secured prior to use. We will continue to provide updates. Please refer to the Customer Portal for the latest information.

User Bulletin for Sequel System: Heat Seal Advisory (Adhesive Seal Warning) ***(NEW!)*** *(May 2017)*

- PacBio advises against the use of adhesive foils and recommends the use of Sequel Sample Plate Foil.
- http://www.pacb.com/wp-content/uploads/User-Bulletin-Heat-Seal-Advisory-Adhesive-Seal-Warning.pdf


Heat Seal Advisory (Adhesive Seal Warning)

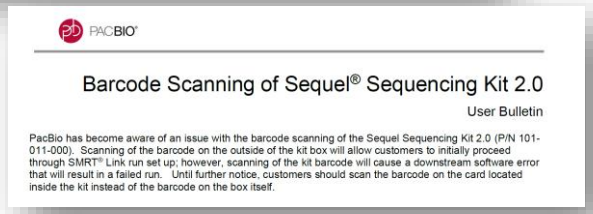
User Bulletin

PacBio has recently become aware of an issue associated with the use of adhesive foil seals. The use of adhesive foil seals can lead to instrument damage. Until further notice, PacBio advises against the use of adhesive foil seals and recommends the of use of the Sequel® Sample Plate Foil (P/N 100-667-400). The of use of the Sequel Sample Plate Foil requires the purchase of a heat sealer. PacBio recommends the Thermo Scientific™ ALPS™ 50 V Semi automated Microplate Heat Sealer (P/N AB-1443).

The use of adhesive foil seals will void existing instrument warranties. Customers will be responsible for any damage resulting from the use of products discordant with official PacBio recommendations.

User Bulletin for Sequel System: Barcode Scanning of Sequel Sequencing Kit 2.0 ***(NEW!)*** *(May 2017)*

- PacBio is providing clarity on which barcode to scan to ensure the Sequel System has the correct information and that all the consumables are compatible.
- http://www.pacb.com/wp-content/uploads/User-Bulletin-Barcode-Scanning-of-Sequel-Sequencing-Kit-2.0.pdf


Barcode Scanning of Sequel® Sequencing Kit 2.0

User Bulletin

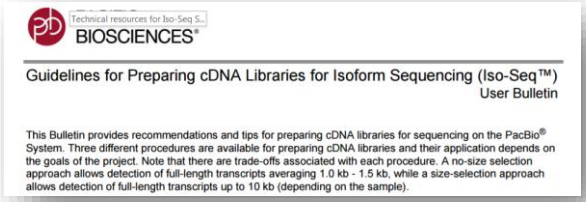
PacBio has become aware of an issue with the barcode scanning of the Sequel Sequencing Kit 2.0 (P/N 101-011-000). Scanning of the barcode on the outside of the kit box will allow customers to initially proceed through SMRT® Link run set up; however, scanning of the kit barcode will cause a downstream software error that will result in a failed run. Until further notice, customers should scan the barcode on the card located inside the kit instead of the barcode on the box itself.

Find all protocols at http://www.pacb.com/support/documentation/

User Bulletins (Cont.)

User Bulletin – Guidelines for Preparing cDNA Libraries for Isoform Sequencing

- http://www.pacb.com/wp-content/uploads/2015/09/User-Bulletin-Guidelines-for-Preparing-cDNA-Libraries-for-Isoform-Sequencing-Iso-Seq.pdf



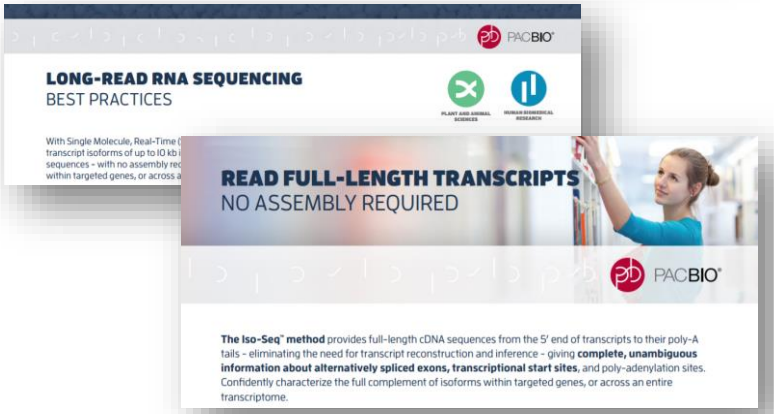
Application Briefs and Brochures

*Application Brief: Long-Read RNA Sequencing Best Practices **(NEW!)** (2017)*

- http://www.pacb.com/wp-content/uploads/Application-Brief-RNA-sequencing-Best-Practices.pdf

Application Brochure: Read full-length transcripts – no assembly required (2016)

- http://www.pacb.com/wp-content/uploads/2015/09/Read-Full-Length-Transcripts-No-Assembly-Required.pdf



Webinars

*Tutorial: Iso-Seq Analysis Application **(NEW!)** (Nov. 2016)*

- http://www.pacb.com/wp-content/uploads/SMRTLink-Video-5-Iso-Seq.mp4

*PAG 2017 PacBio Workshop: Using Iso-Seq to Fill in Your A_no__tion (Richard Kuo, Roslin Institute, University of Edinburgh) **(NEW!)** (Jan. 2017)*

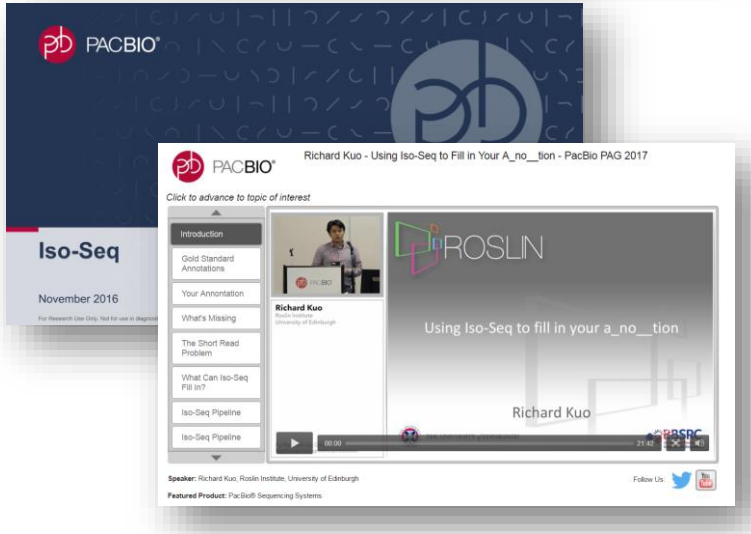
- http://www.pacb.com/wp-content/uploads/SMRTLink-Video-5-Iso-Seq.mp4

PacBio Webinar: Iso-Seq Bioinformatics Webinar (2015)

- http://aa314.gondor.co/webinar/iso-seqtm-analysis-beyond-advanced-bioinformatics-for-transcriptome-sequencing-using-long-reads/

PacBio Webinar: Iso-Seq Sample Preparation and Experimental Design Webinar (2015)

- http://aa314.gondor.co/webinar/iso-seqtm-method-sample-prep-and-experimental-design-for-full-length-cdna-sequencing/



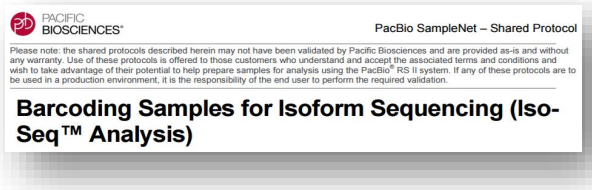
Iso-Seq Library Preparation Protocols

	PacBio RS II	Sequel System (NEW!)
No Size Selection	PACIFIC BIOSCIENCES® Procedure & Checklist - Isoform Sequencing (Iso-Seq™) using the Clontech® SMARTer® PCR cDNA Synthesis Kit and No Size Selection http://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-Isoform-Sequencing-Iso-Seq-using-the-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-and-No-Size-Selection.pdf	PACBIO® Procedure & Checklist - Iso-Seq™ Template Preparation for Sequel™ Systems http://www.pacb.com/wp-content/uploads/Procedure-Checklist-Iso-Seq-Template-Preparation-Sequel-Systems.pdf
With Size Selection	PACIFIC BIOSCIENCES® Procedure & Checklist - Isoform Sequencing (Iso-Seq™) using the Clontech® SMARTer® PCR cDNA Synthesis Kit and Manual Agarose-gel Size Selection http://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-Isoform-Sequencing-Iso-Seq-using-the-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-and-Manual-Agarose-gel-Size-Selection.pdf	
	PACIFIC BIOSCIENCES® Procedure and Checklist - Isoform Sequencing (Iso-Seq™) Using the Clontech® SMARTer® PCR cDNA Synthesis Kit and BluePippin™ Size-Selection System http://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-Isoform-Sequencing-Iso-Seq-using-the-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-and-the-BluePippin-Size-Selection-System.pdf	
	PACBIO® Procedure and Checklist - Isoform Sequencing (Iso-Seq™ Analysis) Using the Clontech SMARTer cDNA Synthesis Kit and SageELF™ Size-selection System http://www.pacb.com/wp-content/uploads/Procedure-Checklist-Isoform-Sequencing-Iso-Seq-Analysis-using-the-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-and-SageELF-Size-Selection-System.pdf	

Multiplexed Iso-Seq Library Preparation Protocols

Shared Protocol – Barcoding Samples for Isoform Sequencing

- http://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Barcoding-Samples-for-Isoform-Sequencing-Iso-Seq-Analysis.pdf


PACIFIC BIOSCIENCES

PacBio SampleNet – Shared Protocol

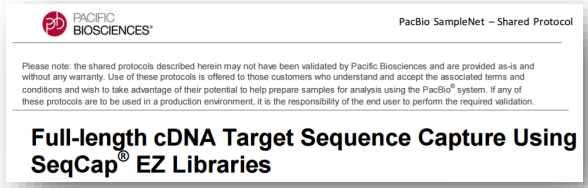
Please note: the shared protocols described herein may not have been validated by Pacific Biosciences and are provided as-is and without any warranty. Use of these protocols is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis using the PacBio® RS II system. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

Barcoding Samples for Isoform Sequencing (Iso-Seq™ Analysis)

Full-length cDNA Target Sequence Capture Protocols

Shared Protocol – Full-length cDNA Target Sequence Capture Using SeqCap® EZ Libraries

- http://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Full-length-cDNA-Target-Sequence-Capture-Using-Roche-NimbleGen-SeqCap-EZ-Library.pdf


PACIFIC BIOSCIENCES

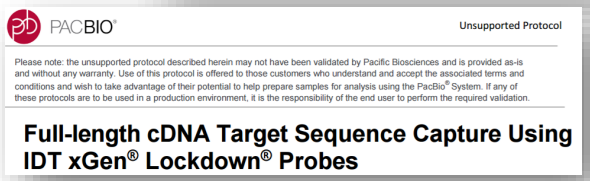
PacBio SampleNet – Shared Protocol

Please note: the shared protocols described herein may not have been validated by Pacific Biosciences and are provided as-is and without any warranty. Use of these protocols is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis using the PacBio® system. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

Full-length cDNA Target Sequence Capture Using SeqCap® EZ Libraries

Unsupported Protocol – Full-length cDNA Target Sequence Capture Using IDT xGen Lockdown Probes

- http://www.pacb.com/wp-content/uploads/Unsupported-Protocol-Full-length-cDNA-Target-Sequence-Capture-IDT-xGen-Lockdown-Probes.pdf


PACBIO

Unsupported Protocol

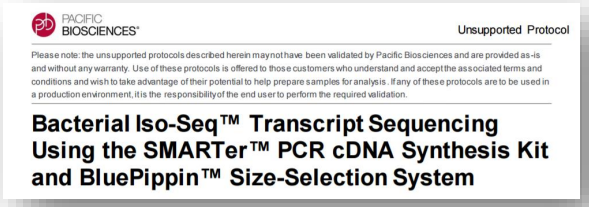
Please note: the unsupported protocol described herein may not have been validated by Pacific Biosciences and is provided as-is and without any warranty. Use of this protocol is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis using the PacBio® System. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

Full-length cDNA Target Sequence Capture Using IDT xGen® Lockdown® Probes

Bacterial Iso-Seq Library Preparation Protocols

Unsupported Protocol – Bacterial Iso-Seq Transcript Sequencing Using the SMARTer PCR cDNA Synthesis Kit and BluePippin Size-Selection System

- http://www.pacb.com/wp-content/uploads/Unsupported-Protocol-Bacterial-Iso_Seq-Clontech-SMARTer-PCR-cDNA-Synthesis-Kit-BluePippin-Size-Selection.pdf


PACIFIC BIOSCIENCES

Unsupported Protocol

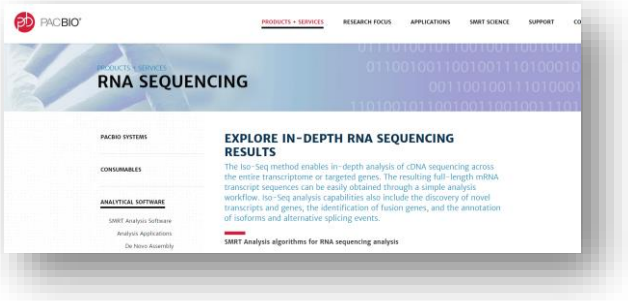
Please note: the unsupported protocols described herein may not have been validated by Pacific Biosciences and are provided as-is and without any warranty. Use of these protocols is offered to those customers who understand and accept the associated terms and conditions and wish to take advantage of their potential to help prepare samples for analysis. If any of these protocols are to be used in a production environment, it is the responsibility of the end user to perform the required validation.

Bacterial Iso-Seq™ Transcript Sequencing Using the SMARTer™ PCR cDNA Synthesis Kit and BluePippin™ Size-Selection System

Find all protocols at http://www.pacb.com/support/documentation/

Iso-Seq Data Analysis

http://www.pacb.com/products-and-services/analytical-software/smrt-analysis/analysis-applications/rna-sequencing/



SMRT Analysis Iso-Seq Bioinformatics Tutorials – PacBio RS II

- https://github.com/PacificBiosciences/cDNA_primer/wiki

SMRT Analysis Iso-Seq Bioinformatics Tutorials – Sequel System

- https://github.com/PacificBiosciences/IsoSeq_SA3nUP/wiki

Iso-Seq Analysis Bioinformatics FAQ

- https://github.com/PacificBiosciences/cDNA_primer/wiki/Iso%E2%80%90Seq-FAQ

PacBio In-House Dataset Releases

https://github.com/PacificBiosciences/IsoSeq_SA3nUP/wiki/Iso-Seq-in-house-datasets

*Human Alzheimer Brain **(NEW!)** (released Sept 2016)*

- http://www.pacb.com/blog/data-release-alzheimer-brain-isoform-sequencing-iso-seq-dataset/

MCF-7 Human Breast Cancer Cell Line (two releases, most recent in May 2015)

- https://github.com/PacificBiosciences/DevNet/wiki/IsoSeq-Human-MCF7-Transcriptome

Human Brain, Heart, and Liver (released Oct 2014)

- http://blog.pacificbiosciences.com/2014/10/data-release-whole-human-transcriptome.html

WHERE TO FIND SMRT RESOURCES

http://www.pacb.com/smrt-science/smrt-resources/

Explore our collection of resources and learn how scientists use SMRT Sequencing to advance their research.

Scientific publications

Explore our database of scientific publications featuring PacBio data.

Conference proceedings

Access conference posters and presentations our customers, collaborators, and internal scientists have presented at various scientific meetings.

PacBio literature

View case studies, brochures, application notes, and more.

Video gallery

Watch our collection of videos, webinars, customer testimonials, and more.

Blog

Read our blog featuring new research, publications, conference summaries, and SMRT Sequencing updates.

Product documentation and training

Visit user documentation for our entire documentation library and training for user training materials.

SUMMARY

- Prepare full-length transcripts using the Clontech® SMARTer® PCR cDNA Synthesis Kit with as little as 1 ng of poly A+ RNA or 2 ng of total RNA
- Sequel System loading protocols reduce need for size selection for transcripts <4 kb
  - Optional size-selection protocols to enrich for transcripts >4 kb
- Survey transcriptomes in 1–2 SMRT Cells on the Sequel System
  - Increase sequencing depth for more comprehensive transcriptome characterization
- Compatible with standard target enrichment methods, such as NimbleGen SeqCap EZ or IDT xGen Lockdown Probes
- Multiplex transcripts or full transcriptomes with sample barcoding
  - Profile transcripts from multiplexed samples in a single Sequel SMRT Cell 1M
- Data analysis protocols and tools available through SMRT Analysis and PacBio DevNet to generate high-quality, full-length transcript sequences with no assembly required
  - Run Iso-Seq analysis in either *de novo* (no genome reference required) or reference-based mode
  - Run Iso-Seq with Mapping analysis (map isoforms to GMAP) to enable studying gene families, gene fusion, accurate identification of unique isoforms.

PACBIO®

PacBio Scientific Conference Poster Presentations

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

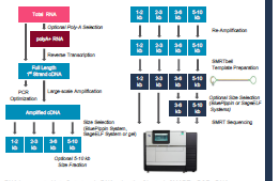
The protein coding potential of most plant and animal genomes is dramatically increased via alternative splicing. Identification and annotation of expressed mRNA isoforms is critical to the understanding of these complex organisms. While microarrays and other NGS-based methods have become useful for 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.

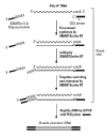
Iso-Seq Sample Preparation Methods

Iso-Seq Sample Preparation Workflow on the PacBio RS II

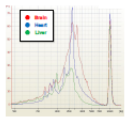


RNA is converted into first strand cDNA using the Clontech SMARTer PCR cDNA Synthesis Kit followed by universal amplification. Amplified cDNA is size fractionated and converted into SMRTbell templates for sequencing on the PacBio RS II.

Clontech SMARTer PCR cDNA Synthesis Kit

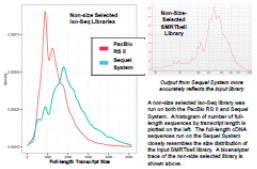


Size Distribution of Amplified cDNA from Multiple Tissues



Optimizing the Iso-Seq Application on the Sequel Platform

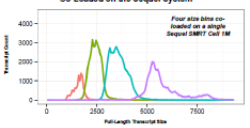
Magbead-loaded Samples on Sequel System Have Decreased Loading Bias



Output from Sequel System more accurately reflects the input library

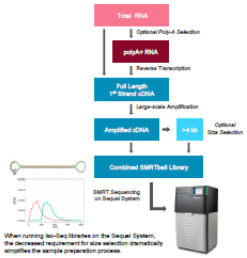
A non-size selected Iso-Seq library was run on both the PacBio RS II and Sequel System. A histogram of number of full-length sequences by transcript length is plotted on the left. The full-length cDNA sequences run on the Sequel System accurately reflect the size distribution of the input SMRTbell library.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

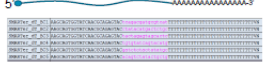
Streamlined Workflow for Iso-Seq Sample Preparation on the Sequel System



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

Barcoding cDNA Libraries

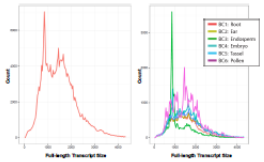
Barcoding During Reverse Transcription Step



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.

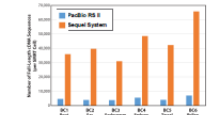
Analyzing the Maize Transcriptome

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



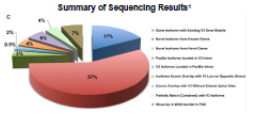
Distribution of full-length transcript sizes across all six tissues combined (left) or each tissue individually (right) using barcodes to segregate the samples. Libraries were non-size selected and run on the Sequel instrument.

Barcode Distribution

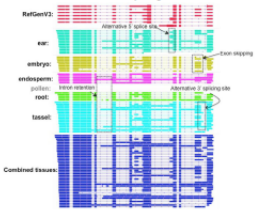


Count of each identified barcode for a single SMRT Cell run on either the PacBio RS II or Sequel instrument.

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.

http://www.pacb.com/wp-content/uploads/Clark-PAG-2017-Full-Length-cDNA-Sequencing-on-the-PacBio-Sequel_Platform.pdf

PAG 2017

A Method for the Identification of Variants in Alzheimer's Disease Candidate Genes and Transcripts Using Hybridization Capture Combined with Long-Read Sequencing

Steve Kujawa[1], Jenny Ekholm[1], Kevin Eng[1], Ting Hon[1], Elizabeth Tseng[1], Aaron Wenger[1], Kristina Giorda[2], Jiashi Wang[2] & Mirna Jarosz[2]
[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025
[2]Integrated DNA Technologies, 1710 Commercial Park, Coralville, IA, 52241

Introduction

Alzheimer's disease (AD) is a devastating neurodegenerative disease that is genetically complex. Although great progress has been made in identifying fully penetrant mutations in genes that cause early-onset AD, these still represent a very small percentage of AD cases. Large-scale, genome-wide association studies (GWAS) have identified at least 20 additional genetic risk loci for the more common form of late-onset AD. However, the identified SNPs are typically not the actual risk variants, but are in linkage disequilibrium with the presumed causative variants[1].

Long-read sequencing together with hybrid-capture targeting technologies provides a powerful combination to target candidate genes/transcripts of interest. Here we present a method for capturing genomic DNA (gDNA) and cDNA from two AD subjects using a panel of probes targeting 35 AD candidate genes. By combining xGen® Lockdown® probes with SMRT Sequencing, we provide completely sequenced candidate genes as well as their corresponding full-length transcripts. Furthermore, we are able to take advantage of heterozygous variants to phase the genes and their corresponding transcript isoforms into their respective haplotypes.

Materials and Methods

A custom panel of 35 AD genes (Table 1) was designed using IDT xGen Lockdown probes. Probes were placed approximately every 1 kb (Figure 1) and designed to cover the entire gene (exons, introns and regulatory regions).

Genes Included in the Panel				
ABCA7	*APH1*	*APOE*	*APP*	*BACE1*
BIN1	*BSG*	*CASS4*	*CD2AP*	*CD33*
CELF1	*CLU*	*CR1*	*EPHA1*	*FERMT2*
GRN	*HLA-DRB1*	*HLA-DRB5*	*INPP5D*	*MAPT*
MEF2C-AS1	*MS4A6A*	*NCSTN*	*NME8*	*PICALM*
PSEN1	*PSEN2*	*PTK2B*	*RIN3*	*SLC24A4*
SNCA	*SORL1*	*TOMM40*	*TREM2*	*ZCWPW1*

Table 1. The custom AD panel includes 35 genes.

Figure 1. Probe design for *PSEN1*. 77 probes were evenly spaced across the ~90 kb gene.

Two subjects were sequenced during this experiment (Table 2). For each subject, gDNA was captured with the custom AD panel according to the published protocol[2] and sequenced on eight PacBio RS II SMRT Cells. Separately, for each subject, RNA was converted to cDNA, captured with the custom AD panel according to the published protocol[3] and sequenced on four PacBio RS II SMRT Cells.

Subject		Source of Genomic DNA	Source of Total RNA
#1	87 year-old male	Brain, Frontal Lobe	Brain, Temporal Lobe
#2	93 year-old female	Skeletal Muscle	Brain, Temporal Lobe

Table 2. gDNA and total RNA from two AD subjects were purchased from BioChain Institute, Inc.

Results - Genes

Reads from the gDNA from Subjects 1 and 2 were mapped to the hg38 reference genome using NGM-LR. Structural variants >50 bp were called using PBHoney Spots (Table 3).

	# Events	# Unique Genes
Deletions >50 bp	15	10
Insertions >50 bp	16	8

Table 3. SVs >50 bp Observed in the 35 AD Genes from Subjects 1 & 2. 31 unique SVs were observed, ranging in size from 65 bp to multiple kilobases.

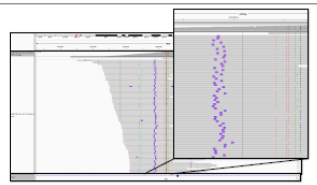


Figure 2. gDNA of *RIN3* gene from Subject 2. Approximately 50 bp insertion (purple bars) found in intron 4 of the *RIN3* gene.

Figure 3. gDNA of *APP* gene from Subject 1. Approximately 550 bp inversion in intron 6 of the *APP* gene.

Results - Transcripts

The captured cDNA from Subjects 1 and 2 were run through the Iso-Seq (ToFU) bioinformatics pipeline to obtain Quiver-polished, full-length, high-quality transcript sequences. Sequences were then mapped to the hg38 genome and filtered with criteria: (1) alignment coverage ≥99%; (2) alignment identity ≥95; (3) at least 5 FL read support; (4) is not a 5' degraded product; and (5) overlaps the probe target region. This resulted in a total of 515 isoforms from Subject 1 and 507 isoforms from Subject 2. To compare with existing annotation, we selected all Gencode v25 transcripts from the target genes with an annotated transcript support level of 1 (most reliable annotation, all junctions supported by at least one mRNA evidence), resulting in 111 isoforms.

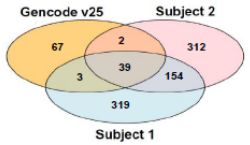


Figure 4. Comparison of isoforms observed in Subjects 1 & 2 with Level 1 isoforms in Gencode v25.

Figure 5. Haplotyped *MAPT* transcripts from Subject 1. Heterozygous SNPs can be used to haplotype the transcripts. A novel exon (red arrows) was observed in three of the five isoforms in Phase 1 and not observed in any of the 21 isoforms in Phase 2.

Results – Haplotyped Variants

After alignment to the hg38 genome, heterozygous variants can be used to further assign the gDNA and transcripts to their appropriate haplotype. As the average fragment size of the captured gDNA is ~6 kb, it is possible to phase regions that are multiple, tens of kilobases in length. Full-length transcripts are easily phased if a heterozygous SNP is captured in an exon or retained intron.

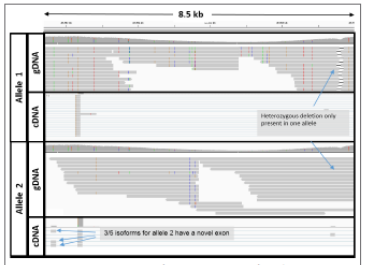


Figure 6. Phased Genes & Transcripts of *MAPT* from Subject 1. Heterozygous SNPs can be used to phase the genomic DNA and transcripts to their appropriate haplotype. Once phased, variants such as this 100 bp heterozygous deletion (blue arrows upper right) can be studied to better understand their potential impact on transcript isoform production. Five unique isoforms were observed from allele 2. Three of these isoforms contained a novel exon (blue arrows lower left) that was only present in allele 2. These exons were flanked by the canonical "AG" and "GT" splice sites in the gDNA.

Conclusion

Combining xGen Lockdown probes with SMRT Sequencing provides a method for completely sequenced candidate genes and their corresponding full-length transcripts.

This method enables:

- Detection of a broad range of genomic variants, from SNPs to multi-kilobase insertions and deletions
- Detection of novel transcript isoforms, including novel exons
- Assignment of variants and transcripts isoforms to their specific alleles

References

1. Van Cauwenberghe C, et al. (2015). The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med*, 18(5), 421-430.
2. Target Sequence Capture Using IDT Library with PacBio® Barcoded Adapters
3. Full-length cDNA Target Sequence Capture Using IDT xGen® Lockdown® Probes

Acknowledgements
The authors would like to thank everyone who helped generate data for the poster.

http://www.pacb.com/wp-content/uploads/Kujawa-AGBT-2017-Alzheimers-Disease-Candidate-Genes-and-Transcripts-Using-Hybridization-Capture.pdf

AGBT 2017

Simplified Sequencing of Full-Length Isoforms in Cancer on the PacBio Sequel Platform

Meredith Ashby, Ting Hon, Elizabeth Tseng, and Tyson A. Clark
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Abstract

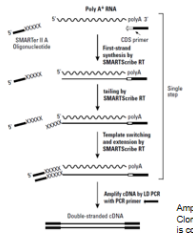
The aberrant transcription and expression of alternative RNA isoforms has been observed in multiple types of cancer, and is hypothesized to contribute to oncogenesis in certain cancer subtypes. Identification and annotation of cancer-specific mRNA isoforms is critical to understanding how mutations in the genome affect the biology of cancer cells. 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. In cancer proteomics studies, the identification of biomarkers from mass spectroscopy data is often limited by incomplete gene isoform expression information to support protein to transcript mapping

The Iso-Seq protocol developed at PacBio offers the only solution for direct sequencing of full-length, single-molecule cDNA sequences needed to discover biomarkers for early detection and cancer stratification, to fully characterize gene fusion events, and to elucidate drug resistance mechanisms. Knowledge of the complete isoform repertoire is key for accurate quantification of isoform abundance. As most transcript sizes range from 1 – 10 kb, fully intact RNA molecules can be sequenced using SMRT Sequencing without requiring fragmentation or post-sequencing assembly. However, some cancer research applications have presented a challenge for the Iso-Seq protocol, due to the combination of limited sample input and the need to deeply sequence heterogenous samples.

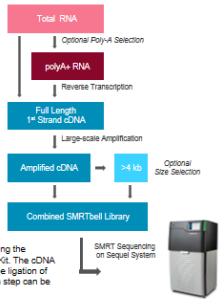
Here, we report the optimization of the Iso-Seq library preparation protocol for the PacBio Sequel platform and its application to cancer cell lines and tumor samples. We demonstrate how loading enhancements on the higher-throughput Sequel instrument have decreased the need for size-fractionation steps, reducing sample input requirements while simultaneously simplifying the sample preparation workflow and increasing the number of full-length transcripts per SMRT Cell. The results highlight the potential for broader application of the Iso-Seq method to more comprehensively characterize alternative splicing in cancer.

Iso-Seq Sample Preparation Methods

Clontech SMARTer PCR cDNA Synthesis Kit



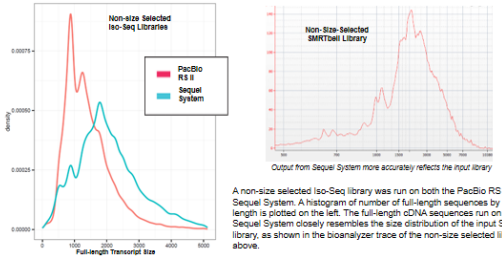
Streamlined Workflow for Iso-Seq Sample Preparation on the Sequel System



Amplified full-length cDNA is generated using the Clontech SMARTer PCR cDNA Synthesis Kit. The cDNA is converted into a SMRTbell library with the ligation of hairpin adapters. An optional size selection step can be included to enrich for larger transcripts.

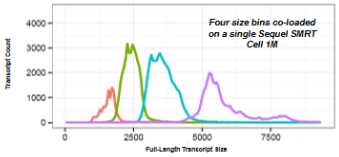
Optimizing the Iso-Seq Application on the Sequel Platform

Magbead-loaded Samples on Sequel System Have Decreased Loading Bias



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, as shown in the bioanalyzer trace of the non-size selected library above.

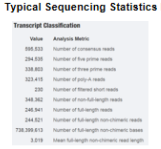
Multiple Size-selected Libraries Can Be Co-loaded on the Sequel System

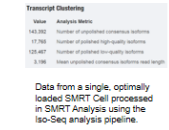


Size-selected Iso-Seq libraries were combined and run together on a single Sequel SMRT Cell 1M. Amplified cDNA had been size fractionated into four size bins (1-2 kb, 2-3 kb, 3-6 kb, and 5-10 kb) using the Sage Science BluePippin system, per the RSII Iso-Seq protocol. Barcoded SMRTbell libraries were made from each of the four size bins, to enable deconvolution post sequencing. A histogram of the full-length transcript lengths from each of the four size bins is shown at left, demonstrating that all four size bins are well represented when co-loaded. To boost representation of the longest isoforms, customers may optionally size select cDNA > 4 kb and pool at a non 1:1 ratio.

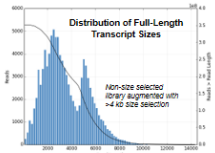
Sequencing of NALM6 Precursor B-ALL Cell Line on Sequel

Typical Sequencing Statistics From a Single Sequel SMRT Cell 1M





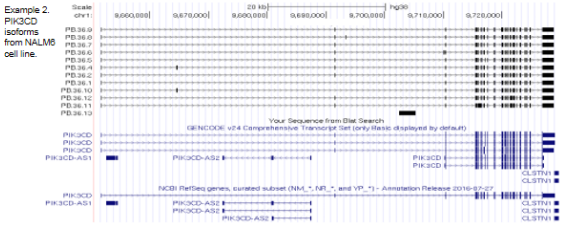
Data from a single, optimally loaded SMRT Cell processed in SMRT Analysis using the Iso-Seq analysis pipeline.



Alternative Splicing in Cancer

Full-Length Transcript Visualization

Example 1. MYC isoforms from NALM6 cell line.



Example 2. PIK3CD isoforms from NALM6 cell line.

Summary and Resources

- PacBio Iso-Seq method generates full-length transcript sequences without the need for assembly of short fragments
- Decreased loading bias on the Sequel System allows for a simplified Iso-Seq sample prep workflow that does not require multiple size-selection steps
- Barcoding during the cDNA generation allows multiplexing samples in a single SMRTbell library
- Improved throughput on the Sequel instrument increases the number of full-length transcripts per SMRT Cell.
- The Iso-Seq method is a powerful tool in the study of cancer providing full-length isoforms, alternative splicing information, and the capability to identify fusion genes.

More information on full-length transcript sequencing (Iso-Seq Application) can be found on the PacBio website: http://pacb.com/isoseq

PDF available upon request

AACR 2017

Poster #687

Full-length cDNA Sequencing of Prokaryotic Transcriptome and Metatranscriptome Samples

Matthew Boitano[1], Bo Yan[2], Ting Hon[1], Elizabeth Tseng[1], Laurence Ettwiller[2] and Tyson A. Clark[1]

[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

[2]New England Biolabs, 240 County Road, Ipswich, MA 01938

PacBio
(650) 521- 6342
mboitano@pacb.com

Introduction

Next-generation sequencing has become a useful tool for studying transcriptomes. However, these methods typically rely on sequencing short fragments of cDNA, then attempting to assemble the pieces into full-length transcripts. Here, we describe a method that uses PacBio long reads to sequence full-length cDNAs from individual transcriptomes and metatranscriptome samples.

We have adapted the PacBio Iso-Seq protocol for use with prokaryotic samples by incorporating RNA polyadenylation and rRNA-depletion steps. In conjunction with SMRT Sequencing, which has average read lengths of 10-18 kb, we are able to sequence entire transcripts, including polycistronic RNAs, in a single read.

Here, we show full-length bacterial transcriptomes with the ability to visualize transcription of operons. We also highlight the ability to detect full-length transcription of operons with alternative start and stop sites. In the area of metatranscriptomics, long reads reveal unambiguous gene sequences without the need for post-sequencing transcript assembly.

Sample Preparation Methods

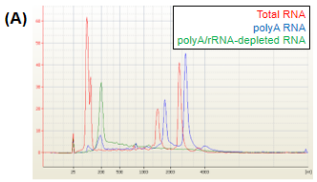
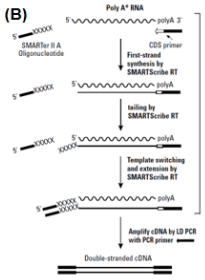
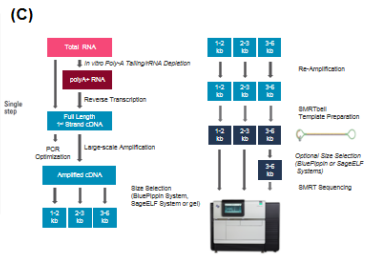


Figure 1. (A) Bioanalyzer traces of *E. coli* total RNA, polyadenylated RNA and polyadenylated/rRNA-depleted RNA. PolyA-tail reaction has been optimized in order to add ~200 nucleotides. Polyadenylated/rRNA-depleted RNA showed good reduction in rRNA peaks and was the input for the cDNA synthesis reaction.

(B) Clontech SMARTer PCR cDNA Synthesis Kit was used to generate double-stranded cDNA. (C) Double-stranded cDNA was size-fractionated using the Sage BluePippin System to sizes of 1-2, 2-3, 3-6 and 5-10 kb (if material is available at each size). This size-fractionated material was converted into SMRTbell libraries. Each library was sequenced on the PacBio RS II with P6C4 chemistry and 4 hour movies. Alternatively, non-size selected material could have been used to generate SMRTbell libraries.





Effects of rRNA Depletion

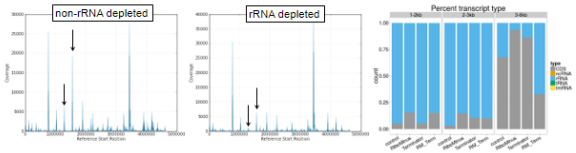


Figure 2. Sequence reads were mapped to the *E. coli* genome reference. Arrows show reduction in coverage of rRNAs after rRNA depletion. Shared peaks are most likely ribosomal associated genes. Out of all traditional rRNA depletion methods tested, RiboMinus had the highest rRNA depletion efficiency.

Detection of Poly-cistronic and Full-length Operon Transcripts

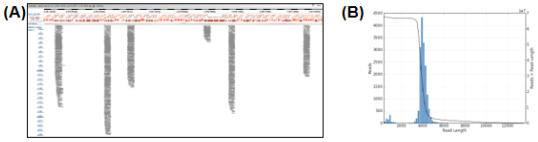


Figure 3. (A) Using long-read SMRT Sequencing, poly-cistronic and full-length operon reads are easily obtained without the need for assembly of short fragments. (B) Data shown are from 3-6 kb size bin, which have an average insert size of 3,917 bp.

Detection of Alternative Transcription Start/Stop Sites

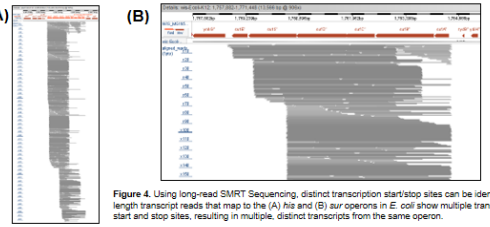


Figure 4. Using long-read SMRT Sequencing, distinct transcription start/stop sites can be identified. Full-length transcript reads that map to the (A) *his* and (B) *sur* operons in *E. coli* show multiple transcription start and stop sites, resulting in multiple, distinct transcripts from the same operon.

Enrich for Primary Transcripts with NEB Cappable-Seq

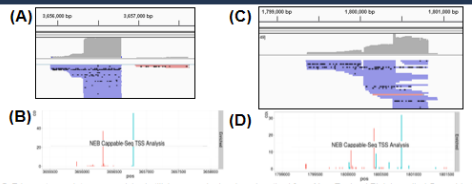


Figure 5. Primary transcripts were enriched utilizing a newly developed method from New England Biolabs called Cappable-Seq (1). Non-processed 5' ends of transcripts were capped with a selectable tag and then sequenced using the SMRT Sequencing Iso-Seq protocol. (A/B) Full-length transcripts, including their transcriptions start sites, were detected allowing for the detection of novel operons. (C/D) This method also provides phasing of transcription start sites and termination sites, even when dealing with overlapping transcripts with additional internal transcription start and termination sites.

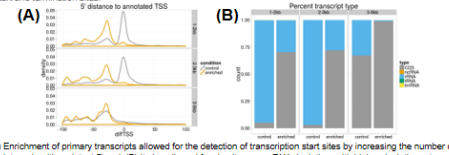


Figure 6. (A) Enrichment of primary transcripts allowed for the detection of transcription start sites by increasing the number of full-length transcript reads with an intact 5' end. (B) It also allowed for simultaneous rRNA depletion, with higher depletion rates compared to other methods.

Metatranscriptome Long-read Sequencing

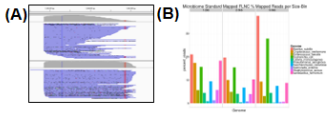


Figure 7. Cells for ZymoBIOMICS Microbial Community Standard were obtained. RNA was isolated using bead beading extraction. RNA was converted to full-length cDNA as described in the Sample Preparation section.

(A) Full-length, polycistronic transcripts were detected, as well as SNV's relative to a reference. These can be used for phasing homologous regions of different organisms.

(B) Percent reads mapped to each genome are dependent on the organism's expression level activity within the community.

Conclusion

- Bacterial SMRT Sequencing Iso-Seq protocol available on PacBio website (2)
- Newly developed method from New England Biolabs allows for enrichment of primary transcripts
- When combining SMRT Sequencing and Cappable-Seq users will be able to detect full-length, polycistronic transcripts from prokaryote transcriptome and metatranscriptome samples. Users will also be able to detect and phase transcription start and termination sites.

References

[1]Ettwiller L., *et al.* (2016) A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* 17, 199.

*http://www.pacb.com/wp-content/uploads/Unsupported-Protocol-Bacterial-Iso_Seq-Clontech-SMARTer-PCR-cDNA-Synthesis-kit-BluePippin-Size-Selection.pdf

PDF Available upon request

ASM 2017

Q&A and Open Discussion

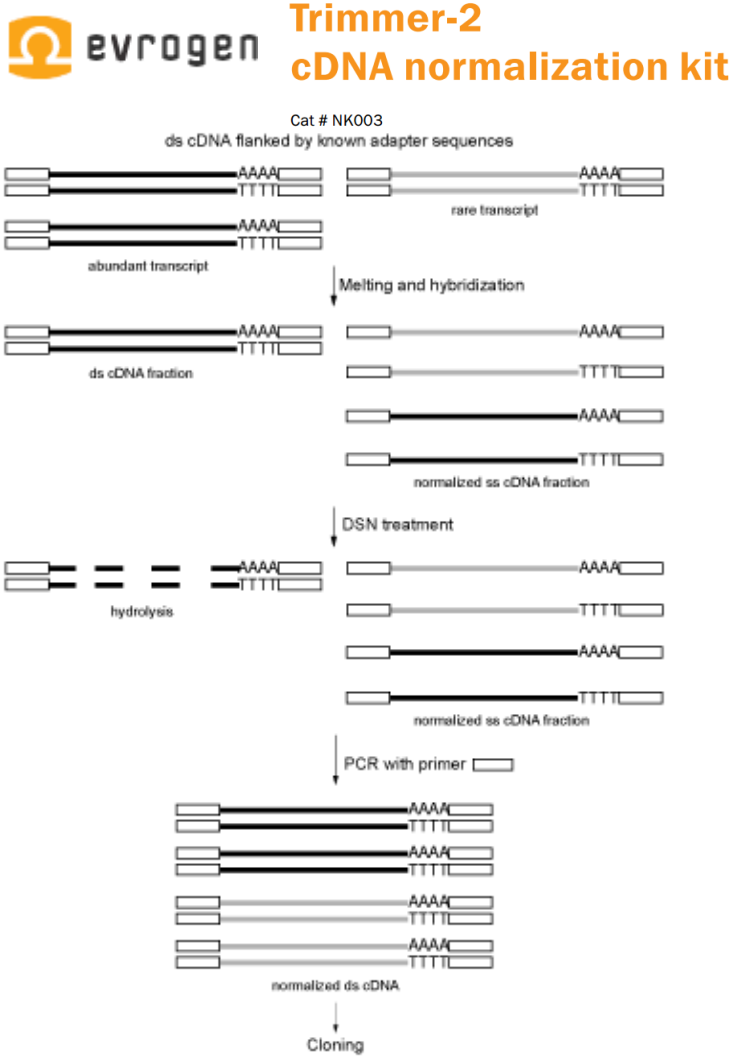
Q&A AND OPEN DISCUSSION

Frequently Asked Questions – Iso-Seq Analysis

Q: ***Does PacBio have any formal recommendations for normalization?***

A: No formal recommendations currently available. PacBio is currently working on generating some recommendations for Normalization on the Sequel System using the Trimmer-2 cDNA Normalization Kit from Evrogen. Unless there is a compelling reason to use Normalization, PacBio does not highly recommend using it yet because we still do not have a complete understanding of the potential impacts to data quality. We do know that longer transcripts can be lost and there is also the possibility of having some transcripts drop out in coverage. If customers are keen to do it, then PacBio can provide some preliminary recommendations – but we are still working to further investigate the utility of performing Normalization with Sequel System Iso-Seq Analysis.

- Normalization reduces the representation of highly expressed genes
- Increases the diversity on a per-sequence basis
- Potential Issues:
  - Transcripts with secondary structure may be degraded
  - Long transcripts may be preferentially removed
  - Rare isoforms of an abundant gene may be lost
- Further work to better understand these methods are ongoing



Q: ***Does PacBio have any formal recommendations for 5' Cap selection?***

A: No formal recommendations currently available. Lexogen makes a kit to do it (https://www.lexogen.com/teloprime-full-length-cdna-amplification/), but PacBio has not fully evaluated it. Early on when we tested it, results were not particularly good. They have updated their kit since then, and we've offered to work with Lexogen to test it, but they haven't responded. It's a good idea in theory, but in practice may have issues (we lost long transcripts, for example). The best bet is to ensure high quality RNA samples and use the standard protocol. However, if customers are keen to do it, the Lexogen kit can be compatible with SMRT Sequencing. We don't have any current recommendations otherwise.

Q: ***Is there an upper limit to the number of barcodes that can be used in multiplexed Iso-Seq analysis?***

A: There is no theoretical limit to the number of barcodes one could use with Iso-Seq analysis, but we've typically recommended that customers stay in the range of 4 – 12 multiplexed samples per pool. The more barcodes you add, the harder it becomes to generate nice, even coverage across all barcodes. If you have to do extra sequencing to improve coverage of one poor-performing barcode, it reduces the overall cost advantages.

Q: ***What are typical secondary analysis metrics like for Sequel Iso-Seq analysis?***

A: Example data from a single, optimally loaded SMRT Cell processed in SMRT Analysis using the Iso-Seq analysis pipeline:

Transcript Classification

Value	Analysis Metric
595,533	Number of consensus reads
294,535	Number of five prime reads
338,803	Number of three prime reads
323,415	Number of poly-A reads
230	Number of filtered short reads
348,362	Number of non-full-length reads
246,941	Number of full-length reads
244,521	Number of full-length non-chimeric reads
738,399,613	Number of full-length non-chimeric bases
3,019	Mean full-length non-chimeric read length

Transcript Clustering

Value	Analysis Metric
143,392	Number of unpolished consensus isoforms
17,765	Number of polished high-quality isoforms
125,467	Number of polished low-quality isoforms
3,196	Mean unpolished consensus isoforms read length

Q: ***Can Iso-Seq analysis be used to identify ncRNAs?***

A: Yes, if they are long enough (> 300 bp) and have a polyA tail. See GitHub repository: https://bitbucket.org/arrigonialberto/lncrnas-pipeline

Frequently Asked Questions - General

How long can I store my polymerase-bound sample?

- ***PacBio RS II:***
  - PacBio recommends that polymerase-bound samples be stored at 4°C and used within 3 days.
- ***Sequel System:***
  - PacBio recommends that polymerase-bound samples be stored at 4°C and used within 7 days.

How do I dissociate my polymerase-bound sample from MagBeads?

- Dissociating polymerase-bound sample from MagBeads may damage the sample and is not recommended. PacBio recommends binding sample to MagBeads immediately before sequencing and proceeding with sequencing as soon as possible. If a delay between MagBead binding and sequencing is unavoidable, Customers can store the sample in the dark at 4°C, but delaying sequencing will be at the Customer's own risk. If a MagBead sample has already been aliquoted into a sample plate, the sample plate should be sealed upon storage at 4°C. For Sequel samples, the sample plate should be heat-sealed with the Sequel Sample Plate Foil (P/N 100-667-400). For PacBio RS II samples, the sample plate should be temporarily sealed with an adhesive microplate sealing film and then the sealing film should be replaced with the PacBio RS II Sample Plate Septum (P/N 000-882-901) before sequencing.

How long can I store my MagBead bound sample?

- PacBio recommends that MagBead samples be stored at 4°C in the dark and sequenced as soon as possible.

My MagBeads were accidentally left at room temperature for several hours. Can they still be used?

- In most cases, MagBeads should still be useable by first chilling them at 4°C before use.

My MagBeads / AMPure beads were accidentally stored at -20℃. Is it still okay to use the beads?

- PacBio does not recommend using AMPure PB beads or MagBeads that have been accidentally stored at -20°C because the beads may become damaged and may leach after being frozen. However, Customers *may* use them at their own risk after bringing the MagBeads to 4°C and AMPure PB beads to room temperature.

When preparing >30 kb SMRTbell libraries, can (AMPure-purified and concentrated) sheared gDNA be stored at 4℃ for longer than 24 hours?

- PacBio generally recommends that AMPure-purified and concentrated sheared gDNA be stored for up to 24 hours at 4°C or at -20°C for longer durations. However, if the gDNA is relatively pure (i.e., free of endonucleases), it should be acceptable to store the sheared gDNA sample for 2-3 days at 4°C.

Conditions for shearing gDNA to a size that can support producing ≥30 kb libraries must be determined and verified empirically for each sample. When preparing ≥30 kb SMRTbell libraries using Megaruptor, what is the recommended target shear size if the desired size selection lower cutoff is, for example, 15-20 kb, 30 kb, or 40 kb?

- When preparing ≥30 kb SMRTbell libraries using Megaruptor, the recommended target shear size depends on the size selection lower cutoff to be employed. The Table below may be considered a useful starting point; but empirical optimization and accurate size quantitation are essential:

Library Insert Size (kb)	Size Selection Lower Cut (kb)	Target gDNA Shear Size (kb)
30	15 - 20	30
30 - 40	15 - 20	50
40 - 50	30	60
50 - 60	40	75

Where can I find the Plate Map and sequences of all the primers in the Barcoded Universal F/R Primers Plate - 96 (P/N 100-466-100) product and Barcoded Adapter Plate - 96 (P/N 100-466-000) product?

- To obtain the sequences of the primers used in the Barcoded Universal F/R Primers Plate - 96 Kit, please contact your local Field Applications Scientist, or submit your inquiry through the PacBio Customer Portal (http://www.pacbioportal.com/) or email techsupport@pacificbiosciences.com.
- The Barcode Plate Map Diagram can be downloaded from PacBio's Documentation webpage (http://www.pacb.com/support/documentation/) here: http://www.pacb.com/wp-content/uploads/2015/09/User-Bulletin-Barcode-Plate-Mapping.pdf

There is a 'Barcoding - RSII and SMRT Analysis 2.3.0 or older' webpage on GitHub (https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/Barcoding). Where can I find the latest guidance on PacBio Barcoding recommendations for multiplexed sample preparation for Sequel System / SMRT Link v4.0 (or later)?

- The most up to date information on PacBio multiplexing applicable to SMRT Link v4.0 (or later) can be found here: https://github.com/PacificBiosciences/SMRT-Link/wiki/SMRT-Analysis-Barcoding-Primer

Can I use Illumina 8-bp barcode index sequences for preparing multiplexed samples for PacBio sequencing?

- No; PacBio does **not** recommend using Illumina 8-bp barcode index sequences for preparing multiplexed samples for PacBio SMRT sequencing applications.

How are the 16-bp PacBio barcodes incorporated into the SMRTbell DNA template?

- PacBio uses two approaches:
  - Adding a barcode to end of the standard SMRTbell adapter. The combined adapter is called a Barcoded Adapter.
  - Adding a barcode to the PCR amplicon. This approach involves a two-step PCR reaction workflow. The internal primers for the first PCR are augmented at the 5' end by universal sequences to the target-specific primers. The external primers contain the 16bp barcode at the 5' end and the universal sequences. This approach is called Barcoded Universal Primers.

What are the supported applications for using PacBio Barcoded Adapters and PacBio Barcoded Universal Primers with multiplexed samples? What are not supported applications?

- Supported applications are sequencing of **one species per sample or loci**. Examples of supported applications include: Confirmation of SNPs, resequencing, most Long Amplicon Analysis (LAA) applications, and Sanger sequencing replacement. An exception is HLA typing, which may have 2 species per loci. Multiplexing of HLA has also been demonstrated with the use of additional custom analyses (see PacBio's AGBT 2015 Poster: http://files.pacb.com/pdf/Poster_MultiplexingHumanHLAGenotyping_DNABarcodeAdapters_HighThroughputResearch.pdf)
- Note: The product specifications for the PacBio Barcoded Adapter Kit and PacBio Barcoded Universal Primer Kit are such that the level of barcode oligo contamination in the 96-plate wells should not exceed 5%. Therefore it is possible, though unlikely, to have 1 other contaminant barcode primer/adapter sequence present at levels up to 5%. PacBio does not recommend using the PacBio Barcoded Adapter Kit and PacBio Barcoded Universal Primer Kit for minor variant detection < 10%.

Does PacBio have any specific DNA polymerase enzyme or Kit recommendations for long-range PCR (LR PCR) for generating long DNA amplicon samples for sequencing?

- While PacBio does not recommend a specific enzyme, a high-fidelity enzyme is generally preferred. For example, PrimeStart GXL from Takara and ThermoFisher Phusion Hot Start II DNA Polymerase have given good results to our internal scientists.

Other Discussion Points

- ***Do these protocols/tools serve well for your RNA sequencing needs?***
- ***What other things would you like us to add to our current solution?***
  - Quality Control (QC) guidelines for Sequel System Iso-Seq Analysis samples
  - More Sequel System Iso-Seq Analysis public dataset releases
- ***What are your opinions on the current state of the application on PacBio SMRT sequencing?***

PACBIO®

www.pacb.com