Guidelines for Preparing cDNA Libraries for Isoform Sequencing (Iso-Seq™)
User Bulletin

This Bulletin provides recommendations and tips for preparing cDNA libraries for sequencing on the PacBio® System. Three different procedures are available for preparing cDNA libraries and their application depends on the goals of the project. Note that there are trade-offs associated with each procedure. A no-size selection approach allows detection of full-length transcripts averaging 1.0 kb - 1.5 kb, while a size-selection approach allows detection of full-length transcripts up to 10 kb (depending on the sample).

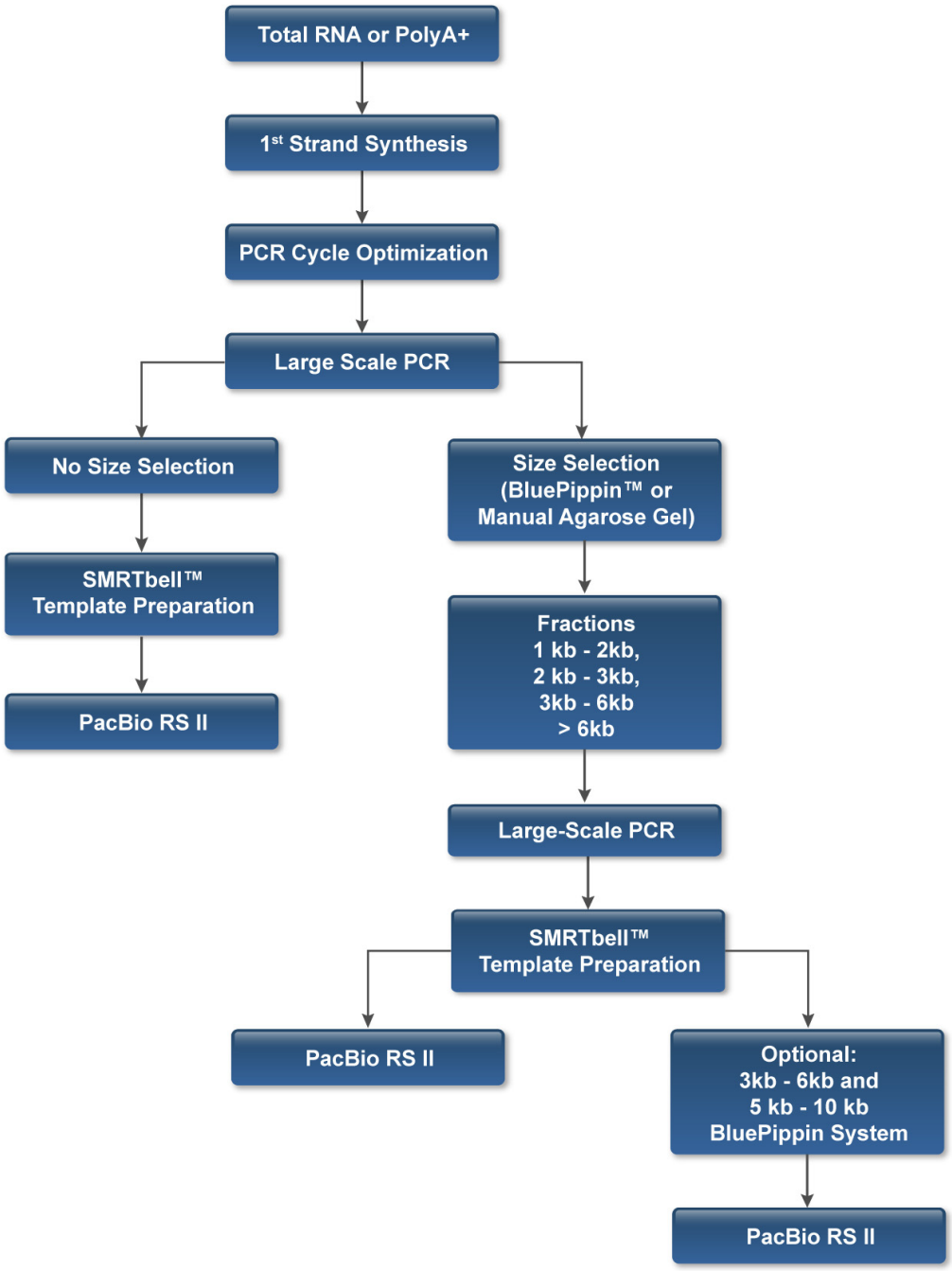

Figure 1 The Library Preparation Workflow

Input RNA Material

The Clontech® SMARTer® PCR cDNA Synthesis Kit can be used with either total RNA or polyA+ RNA. The minimum starting amount is 2 ng of total RNA or 1 ng of polyA+ RNA. The following kits are recommended for polyA+ extraction:

- Ambion® Poly(A) Purist™ Kit http://products.invitrogen.com/ivgn/product/AM1916
- Ambion® Poly(A) Purist™ MAG Kit http://products.invitrogen.com/ivgn/product/AM1922

Note that approximately 1-3% of the total RNA mass will be polyA+ RNA.

cDNA Synthesis using the SMARTer PCR cDNA Synthesis Kit

Clontech's reverse transcriptase (RT) begins synthesis at the polyA+ tail and then synthesizes a cDNA complement to the RNA. Once the SMARTScribe Reverse Transcriptase reaches the 5' end of the mRNA, the enzyme's terminal transferase activity adds a few extra nucleotides to the 3' end of the cDNA.

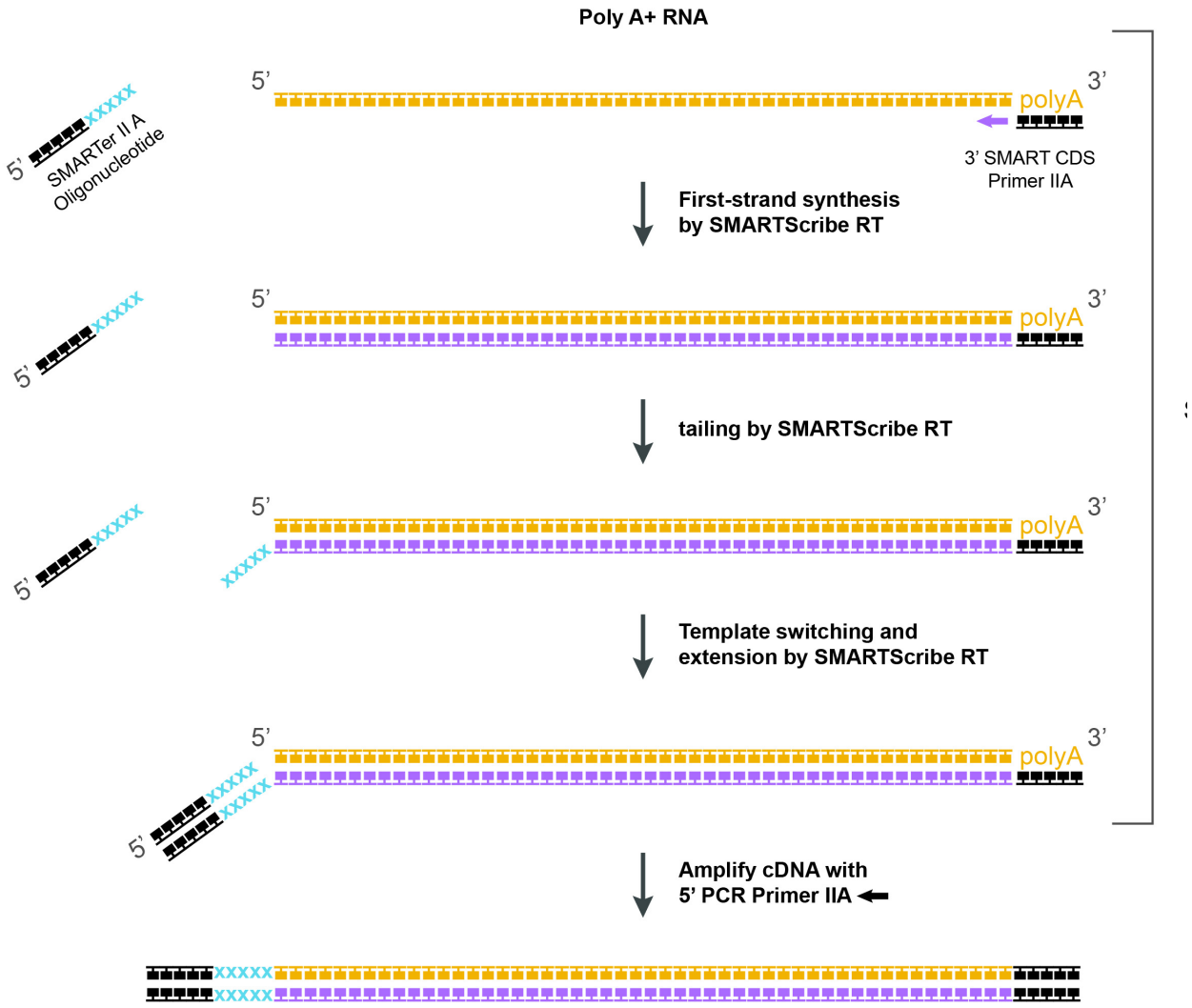

Figure 2 PolyA+ Tail Synthesizes to Double-Stranded cDNA.

The extra nucleotides serve as the binding site for a synthetic oligonucleotide with a complimentary 3' stretch. The RT then switches templates to the synthetic oligo and continues extending the first-strand cDNA to the end of the oligo, creating a universal 3' sequence for second strand synthesis and PCR amplification. The advantage of this mechanism is that a relatively low amount of polyA+ RNA can be used as starting material. Full-length cDNA libraries can be produced from as little as 1 ng of polyA+ RNA, or 2 ng total RNA. The total time from input RNA to SMRTbell™ template library generation is 3 to 4 days.

The terminal transferase activity is only active when the RT enzyme reaches the end of the cDNA molecule, therefore, transcripts that do not synthesize to the end of the 5' mRNA sequence (early termination) will not be converted into templates for second strand synthesis. Partially degraded RNA fragments with polyA+ tails can still be converted to cDNA if the enzyme reaches the end of that degraded sequence. Therefore, high quality RNA is of utmost importance for the detection of long full-length transcripts. The SMARTer PCR cDNA Synthesis Kit does not differentiate between a full-length mRNA sequence and that of a degraded mRNA fragment. If the input RNA is not high quality, the result is a cDNA library that will show an overall 3' bias.

Instructions for first-strand cDNA synthesis, from the Clontech SMARTer PCR cDNA Synthesis Kit, have been integrated into the PacBio supported Iso-Seq document.

Primers for the cDNA Amplification Reaction

PCR amplification of cDNA, produced using the Clontech method, utilizes a single primer (shown below) since both ends of the cDNA share a common sequence:

5' AAG CAG TGG TAT CAA CGC AGA GTA C 3'

However, the primer included in the Clontech kit may not be sufficient for the Iso-Seq procedure. We recommend purchasing additional primers from any oligo synthesis vendor (e.g., IDT, Operon, etc.). Prepare a stock of 12 µM primer and follow the procedure below.

PCR Cycle Optimization

Note that cDNA can be amplified with primers specific to the ends of the cDNA molecules. It is important to optimize the PCR cycle number to ensure adequate double-stranded cDNA for subsequent reactions and size selection while minimizing artifacts from over-amplification. For this purpose, we highly recommend first setting up test amplifications for each cDNA sample using the KAPA™ HiFi High-Fidelity DNA Polymerase from KAPA Biosystems.

Assaying the PCR Optimization by Agarose Gel, Qubit® and/or Bioanalyzer® Systems

Agarose gels can be used for a quick visual inspection of the samples after PCR cycle optimization. Load 5 µL of each sample on a 0.8% agarose gel or on a Lonza FlashGel®.

For a typical eukaryotic cell, the cDNA appears as a smear from 500 bp to approximately 4 kb or 5 kb. At times, discreet bands can be seen within the smear (this is normal) and represent cDNAs from abundant transcripts. There may be a strong band toward the bottom of the gel from primer dimers or free primers. These will be removed during subsequent purifications and will not complicate sample preparation.

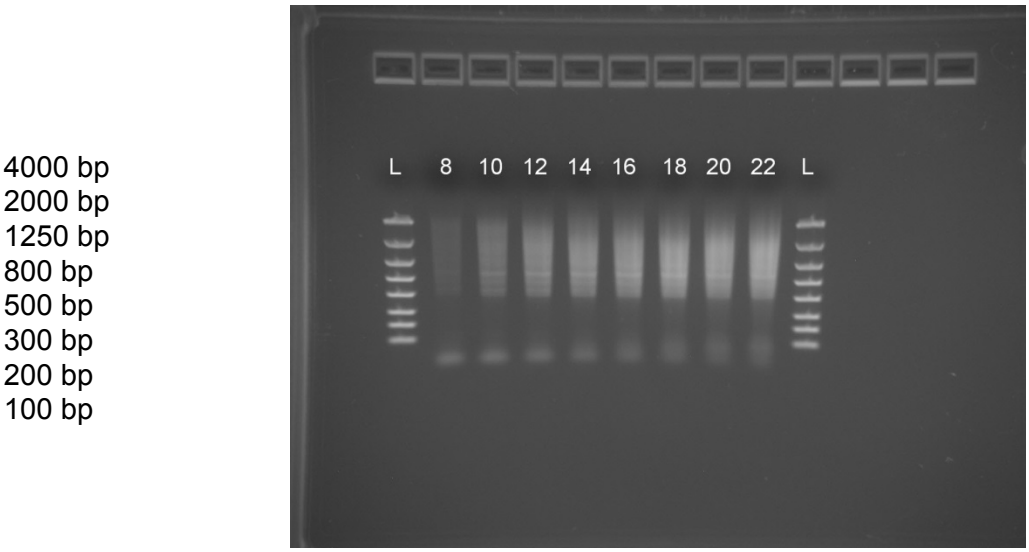

Figure 3 Analysis for optimizing PCR cycles: This sample is Rat Smooth Muscle PolyA+ RNA from Clontech. 100 ng of starting RNA was used to set up cDNA synthesis. Gel electrophoresis was performed on a 1.2% Lonza® FlashGel. Lane L: Lonza FlashGel Ladder (100 bp – 4 kb). The numbers above the other lanes indicate cycle number. 2 reactions of 50 µL PCR were set up to obtain this series cycle of optimization. In this example, PCR with 10 cycles was determined to be optimal for large-scale amplification. Cycles 8 and 10 look similar in banding and size, however cycle 10 has more products than 8 cycles. In cycles 12 and above, the cDNA smears show over-amplification of smaller fragments.

In general, choose the cycle number with the highest yield at the least number of cycles. Avoid conditions that produce extra bands or skew the size distribution to smaller fragments. An example of choosing an optimal PCR cycle is shown in Figure 3.

The Qubit dsDNA quantification system and Bioanalyzer DNA 7500 or 12000 sizing systems can also be used to determine the optimal number of PCR cycles. Following the guidelines above, use the Bioanalyzer system to determine size distribution and the Qubit system to determine quantitation. It should be noted that the Bioanalyzer sizing is very sensitive to the input mass, therefore quantitate using the Qubit or PicoGreen® systems first, and normalize the mass between the wells. Standard measurement of A260 using the Nanodrop® or other spectrophotometers is not recommended, since PCR reactions contain residual primers and nucleotides that can skew the readings.

Non Size-Selected cDNA Library

In general, a non-size-selected cDNA library has a mean insert-size distribution of approximately 1.7 kb and ranges from a few hundred bases to several thousand. However, due to amplification and sequencing bias, the cDNA library size distribution does not translate directly to the length of insert distribution. Loading is biased toward shorter inserts with the majority of full-length transcripts being in the 1 kb – 1.5 kb range (see Figure 4). Depending on the project goals, consider a no size-selection workflow (see the Pacifc Biosciences *Procedure & Checklist - Isoform Sequencing (Iso-Seq™) using the Clontech® SMARTer® PCR cDNA Synthesis Kit and No Size Selection)*.

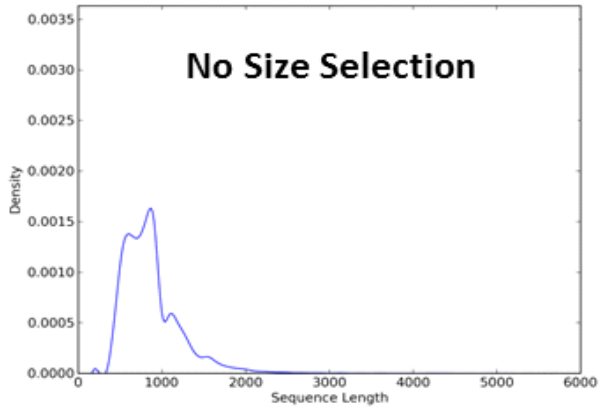

Figure 4 No Size-Selection, Full-Length Transcript Distribution

Size-Selected cDNA Library

To capture a broad range of transcripts, consider collecting 1 kb – 2 kb, 2 kb – 3 kb, 3 kb – 6 kb and > 6 kb fractions. Fractions up to 6 kb can be achieved through either manual gel-cutting, the BluePippin™ system or SageELF™ system. Amplification and sequencing of full-length transcripts > 6 kb can also be achieved based on the sample type, quality or size-selection method. PacBio recommends the BluePippin or the SageELF systems for transcripts > 6 kb. Figure 5 shows full-length transcript distributions generated from 1 kb - 2 kb, 2 kb - 3kb and 3 kb - 6kb fractions size-selected using BluePippin system's 0.75% DF 2 – 6kb Marker S1 cassette definition file. Manual gel-cutting (if performed according to recommendations) is also effective (as previously shown).

With the BluePippin system, the length distribution mode of the 1-2 kb fraction may be noticeably larger than expected (> 2 kb). This can result in loss of < 1.5 kb transcripts. If so, consider constructing a SMRTbell library from a non-size selected sample. This library size should capture transcripts < 1.5 kb. New cassette definition files are currently being evaluated to further improve accuracy of size selections.

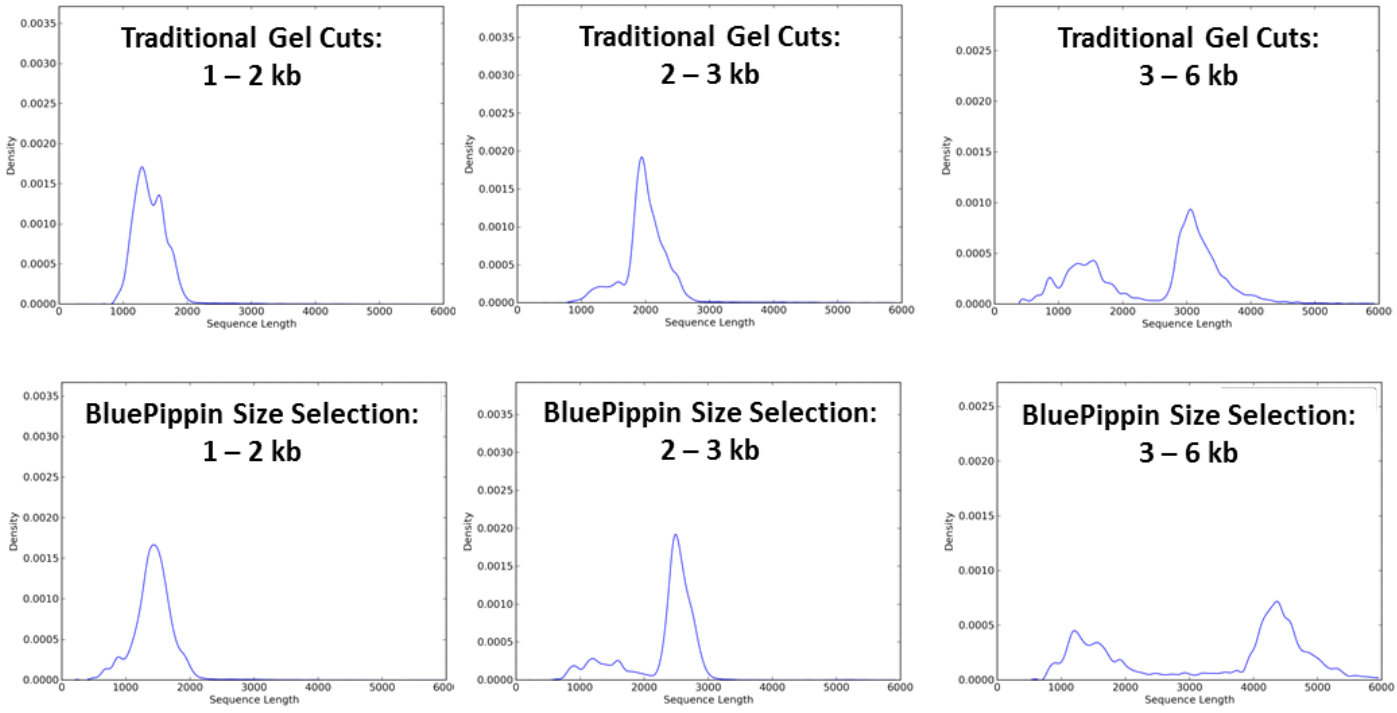

Figure 5 Distribution of sequence lengths of 3 size fractions, using two sizing methods. BluePippin size-selection contains slightly longer transcripts compared to traditional gel-cuts.

Furthermore, the 3 kb - 6 kb fraction will exhibit short insert contamination as shown in figure 5. Note that removal of short inserts (1.5 kb) during size selection (BluePippin and traditional gel cut) is currently not optimal and their presence affects (to some degree) sequencing and detection of long transcripts. In the PacBio system, short SMRTbell templates load more efficiently than longer SMRTbell templates, thereby reducing sequencing and detection of long full-length transcripts. This becomes more noticeable when targeting > 3 kb transcripts.

To maximize sequencing and detection of long transcripts (> 3 kb), consider running the SMRTbell libraries (fractions 3 kb - 6 kb and > 6 kb) in the BluePippin system to remove the small transcripts completely (see Figure 6 below).

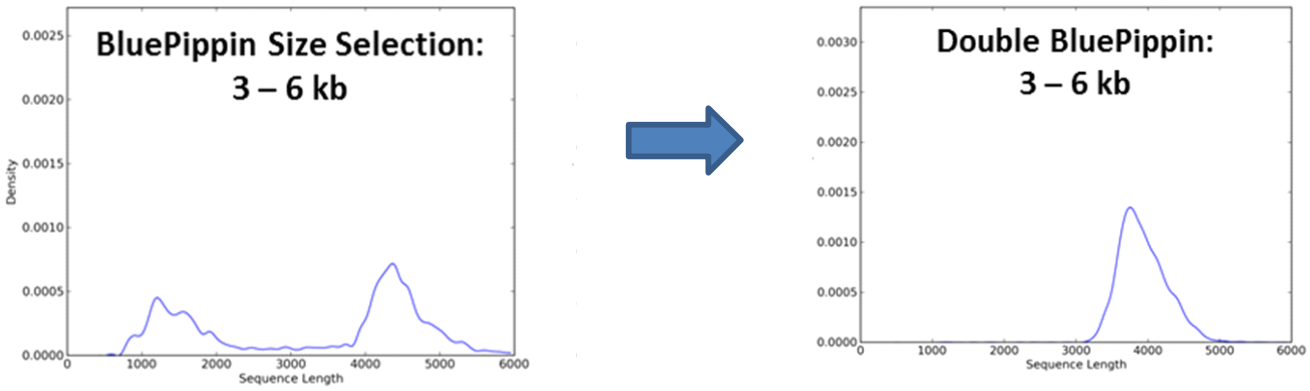

Figure 6 Complete removal of short transcripts can be achieved by running the SMRTbell libraries in the BluePippin system.

Manual Agarose Size Selection

While using the BluePippin system offers automation for size selection, traditional gel cutting (although manual) is also effective. For this method, we have consistently generated good size selected libraries targeting a broad range of transcripts up to 6 kb. However, size selection of > 6 kb has not been evaluated.

To increase success, we list the following guidelines and recommendations for traditional gel cutting.

1. Pour 0.8% agarose gel in 1X TAE buffer.
2. Choose an electrophoresis buffer chamber that allows a slow run.
3. Run gel at 4.2 V/cm for approximately an hour.
    - Measure the distance between the two electrodes (in cm) and multiply by 4.2 to determine the voltage required for running the gel.
4. Add the appropriate amount of dye.
    - We recommend usinng the SYBR® Safe DNA Gel Stain. Add 1 µL to a 10 mL solution.
    - Do not use ethidium bromide with a UV lamp for size selection. Exposure to UV damages the DNA which can result in sub-optimal read lengths.
5. Load the ladder and samples in every other lane (alternating lanes - see Figure 7). Flanking the size ladder on both sides of the sample is highly recommended as it will help in performing gel cuts more precisely. It also mitigates unusual migration issues during electrophoresis.

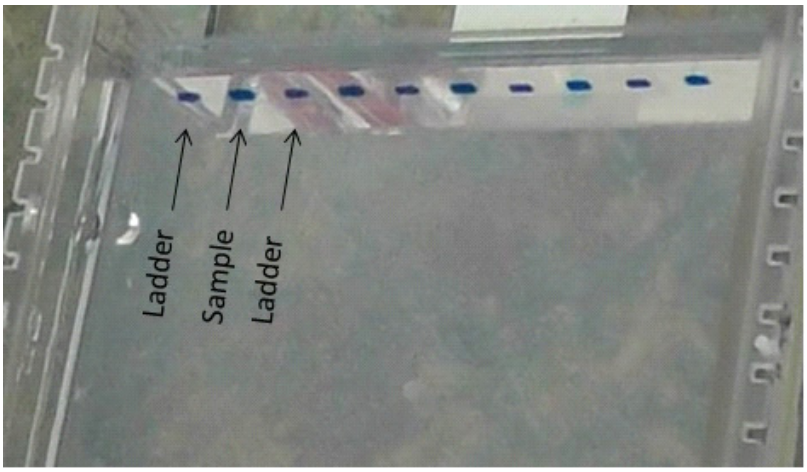

Figure 7 Ladders and Samples Loaded in Alternating Lanes

Depending on sample availability, load the sample in one well or multiple wells (500 ng/well). For samples loaded in one well, cut the gel sections as shown in Figure 8 below.

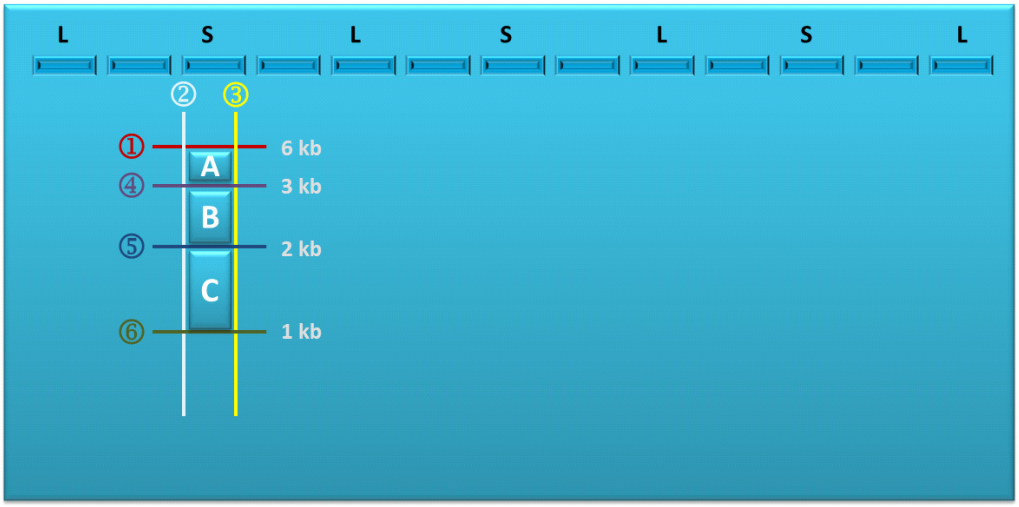

Figure 8 One Well Loading

1. Set aside six clean and unused razor blades.
2. Using clean gloves and a new razor blade, cut 1 first.
3. Discard the razor blade and use a new razor blade to cut 2 next.
4. Repeat step 3 for cuts 3 and 4.
5. Using a sterile small bore pipet tip, stab the gel at an angle so that you are lifting the slice up and toward the wells (away from the smaller fractions). Transfer the gel slice to a fresh 2 mL tube.
6. Using a new razor blade, cut 5.
7. Using a sterile small bore pipet tip, stab the gel at an angle so that you are lifting the gel slice up toward where gel slice A used to be, and transfer the gel slice to a fresh 2 mL tube.
8. Repeat the process for slice C.

Loading samples on separate wells and excising each fraction on a separate lane minimizes short insert contaminations. Consider loading and cutting the gels as shown in Figure 9 below.

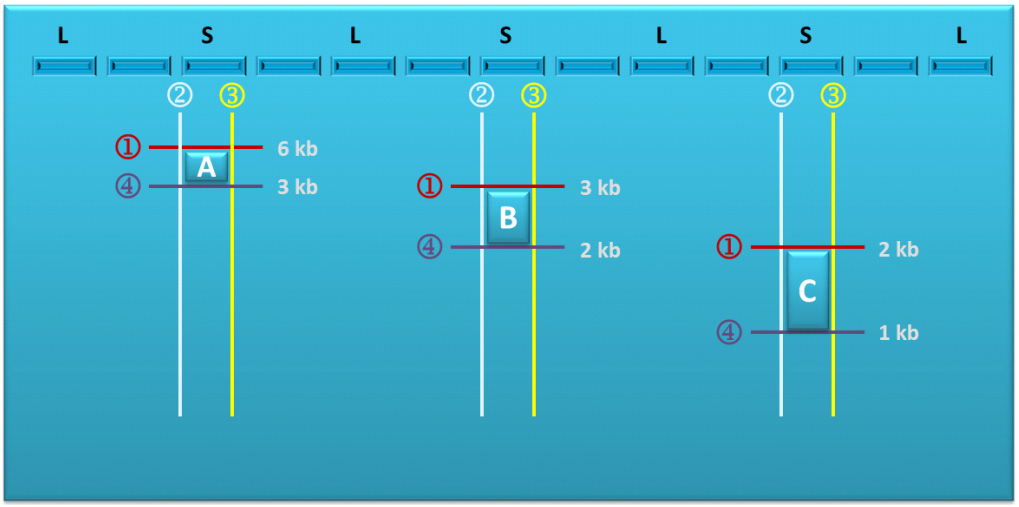

Figure 9 Multiple Lane Loading Minimizes Short Insert Cross Contaminations.

1. Set aside 12 clean unused razor blades.
2. Using clean gloves and a new razor blade, cut 1 first.
3. Discard the razor blade and use a new razor blade to cut 2 next.
4. Repeat step 3 for cuts 3 and 4.
5. Using a sterile small bore pipet tip, stab the gel at an angle so that you are lifting the slice up and toward the wells. Transfer the gel slice out and into a fresh 2 mL tube.
6. Repeat steps 2 to 6 for gel slices B and C.

Anneal and Bind BluePippin™ Size-Selected SMRTbell™ Templates

To anneal sequencing primer and bind polymerase to SMRTbell templates, see the Calculator. For more information about using the calculator, see the Pacific Biosciences *Template Preparation and Sequencing Guide*.

Sequencing Recommendations:

Loading mechanism: MagBead loading

Chemistry: P5 polymerase with C3 Chemistry

Movie Collection: 120 mins

Stage or no-Stage: no difference

On-plate concentration recommendations: Perform titrations

1-2 kb: 25 pM
2-3 kb: 25 pM
3-6 kb: 50 pM
>6 kb: 25-50 pM

MagBead loading is suggested for all fractions and for non-size selected libraries. We recommend performing loading titrations to determine an appropriate loading concentration. The Calculator provides recommended sample concentrations for binding polymerase/template complexes to MagBeads, and for loading complexes on PacBio systems. For information on how to prepare and sequence using MagBeads, see the Pacific Biosciences *Procedure & Checklist - Preparing MagBeads for Sequencing*.

PN 100-378-900-01