Technical Overview: Single-Cell Iso-Seq Library Preparation Using SMRTbell Express Template Prep Kit 2.0

Sequel System ICS v8.0 / Sequel Chemistry 3.0 / SMRT Link v9.0
Sequel II Systems ICS v9.0 / Sequel II Chemistry 2.0 / SMRT Link v9.0
Sequel Ile System ICS v10.0 / Sequel II Chemistry 2.0 / SMRT Link v10.0
Single-Cell Iso-Seq Library Preparation Using SMRTbell Express Template Prep Kit 2.0

1. Single-Cell Iso-Seq Workflow Overview & Experimental Design Considerations
2. Single-Cell Iso-Seq Library Sample Preparation Workflow Details
3. Single-Cell Iso-Seq Library Sequencing Workflow Details
4. Single-Cell Iso-Seq Data Analysis Recommendations
5. Single-Cell Iso-Seq Library Example Sequencing Performance Data
6. Technical Documentation & Applications Support Resources
SINGLE-CELL FULL-LENGTH TRANSCRIPT ISOFORM SEQUENCING (scISO-SEQ METHOD):
HOW TO GET STARTED

**Library Construction, Sequencing & Analysis**

- **Single-Cell cDNA Synthesis**
  Use any third-party single-cell platform to perform cDNA synthesis and amplification

- **Library Construction (SMRTbell Express TPK 2.0)**
  Purify Single-Cell Iso-Seq Library Using ProNex Beads

- **HiFi Sequencing**
  Generate 1,000 unique reads/single cell for up to 3000 cells per SMRT Cell 8M

**Data Analysis**

Single-Cell Iso-Seq Analysis Tools (GitHub)

**Application**

**Application-Specific Best Practices Guide**

**Application-Specific Procedure & Checklist**

**Application Consumable Bundle Purchasing Guide**

**Application Brief: Single-cell RNA sequencing with HiFi reads - Best Practices** (BP109-102020)
Summary overview of application-specific sample preparation and data analysis workflow recommendations

**Procedure & Checklist – Preparing Single-Cell Iso-Seq™ Libraries Using SMRTbell® Express Template Prep Kit 2.0**
Technical documentation containing sample library construction and sequencing preparation protocol details

**PacBio Application Consumable Bundle Purchasing Guide** (PG100-082620)
Purchasing Guide enables users to easily order required consumables needed to prepare a SMRTbell library to run a specific type of application on the Sequel II and IIe Systems*

* Application Consumable Bundles include reagents for library construction, primer annealing and polymerase binding. Core PacBio-branded SMRT Sequencing consumables (SMRT Cells, Sequencing Kits & SMRT Oil), plastics and other 3rd-party reagents are not included in the application bundles.
Single-Cell Iso-Seq Workflow Overview & Experimental Design Considerations
SINGLE-CELL ISO-SEQ SAMPLE PREPARATION PROCEDURE DESCRIPTION


- Single-Cell cDNA samples to be used for this procedure are first generated using a third-party single-cell isolation platform.

  - cDNA products from a typical single-cell preparation are initially re-amplified to increase the mass.

  - Following cDNA re-amplification, SMRTbell Express Template Prep Kit 2.0 is then used for SMRTbell library construction.

APPLICATIONS

RNA SEQUENCING

Single-Cell Full-Length Transcript Isoform Sequencing (scIso-Seq Method)
SINGLE-CELL ISO-SEQ METHOD FEATURES AND EXPERIMENTAL DESIGN CONSIDERATIONS

- **Generate full-length transcript isoforms that can be assigned to individual cells** to characterize alternative splicing and cell heterogeneity within tissues.
  - Iso-Seq method uses single-molecule, real-time (SMRT) sequencing technology to produce highly accurate long reads (HiFi reads) using the Sequel, Sequel II and Sequel IIe Systems.
  - Iso-Seq reads can span the entire 5′ to 3′ end of a transcript – *with no assembly required*.

- Although PacBio does not have a specific single-cell partner or system recommendation, in principle, practically any single-cell platform should be compatible with single-cell Iso-Seq library preparation so long as that platform generates cDNA.
  - For the Iso-Seq method to achieve full-length cDNAs, it is recommended to use a template-switching oligo (TSO). This is a common technique and is currently used in single-cell platforms and PacBio’s current bulk Iso-Seq methods.

- For optimal analytical results, **PacBio recommends combining matching short-read and Iso-Seq datasets** (generated for the same exact single-cell library sample).
  - We recommend that the post-reamplification cDNA yield allow for parallel processing of both short-read sequencing and SMRT Sequencing.

- The **Sequel System requires >80 ng of cDNA**, while the **Sequel II System requires >160 ng cDNA**. These are target DNA amounts for the PCR re-amplification steps for the Iso-Seq library construction workflow (see Page 4 of the procedure).
SINGLE-CELL ISO-SEQ METHOD FEATURES AND EXPERIMENTAL DESIGN CONSIDERATIONS (CONT.)

- cDNA re-amplification is typically achieved by using the **PCR primers specific to a single-cell isolation platform.**

  - If these are not supplied in the quantity required to generate sufficient cDNA input material for both short-read and SMRT Sequencing library construction, **order additional amounts of primers separately.** (PCR primer oligo sequences can be typically obtained from the single-cell platform provider.)

- A sizing platform such as the Agilent Bioanalyzer is a useful tool for checking the quality of the amplified cDNA; the mean size of a high-quality single-cell SMRTbell library is usually **~1.5 kb with molecules that stretch into the 5–6-kb range.**

- The PacBio single-cell sample prep workflow from cDNA re-amplification to SMRTbell Express library preparation takes **approximately 5 hours** to complete.

- Obtain up to ~3 Million full-length reads with cell barcode and UMI information per Sequel II SMRT Cell 8M

  - 1,000 unique transcript reads / single-cell library prep using 3000 cells input
  - 10,000 unique transcript reads / single cell library prep using 300 cells input
Pacific Biosciences does not sell a kit for carrying out the Single-Cell RNA Sequencing method. Use of these methods may require rights to third-party owned intellectual property.

Single-Cell cDNA Preparation
(Third-Party Single-Cell Isolation System)

1. **Poly(A)** mRNA
2. **Reverse transcription (1st-strand synthesis) and addition of 'C' bases at transcript 5' end**
3. **Template switching (from mRNA to TSO) and completion of transcript extension**
4. **2nd-strand synthesis and cDNA amplification**
5. **cDNA Cleanup**
6. **Full-length barcoded single-cell cDNA for input into SMRTbell library construction**

---

Single-Cell SMRTbell Library Preparation
(PacBio Sequel, Sequel II and Sequel Ile Systems)

1. **DNA Damage Repair / End Repair / A-Tailing**
2. **Adapter Ligation**
3. **SMRTbell Purification**
4. **SMRTbell Purification**
5. **Template Switch Oligo (TSO)**
6. **RT-PCR Primer**
7. **Forward PCR Primer**
8. **Reverse PCR Primer**

---

† Pacific Biosciences does not sell a kit for carrying out the Single-Cell RNA Sequencing method. Use of these methods may require rights to third-party owned intellectual property.
Single-Cell Iso-Seq Library Sample Preparation Workflow Details
PROCEDURE & CHECKLIST – PREPARING SINGLE-CELL ISO-SEQ LIBRARIES USING SMRTBELL EXPRESS TEMPLATE PREP KIT 2.0

- This document (PN 101-892-000) contains instructions for constructing Single-Cell Iso-Seq SMRTbell libraries for sequencing on the Sequel, Sequel II and Sequel IIe Systems (Sequel Systems).

- Two-step Single-Cell Iso-Seq sample prep workflow:
  - Intact (un-sheared) RT-PCR products initially generated using a third-party single-cell preparation system are first re-amplified to increase the mass.
  - SMRTbell Express Template Prep Kit 2.0 is then used for SMRTbell library preparation.

- Protocol document contains:
  1. General laboratory best practices and input RNA QC recommendations
  2. Instructions for performing single-cell cDNA re-amplification prior to SMRTbell library construction
  3. Instructions for constructing SMRTbell libraries using amplified single-cell cDNA products and SMRTbell Express Template Prep Kit 2.0
  4. Sample setup guidance for preparing Single-Cell Iso-Seq SMRTbell libraries for sequencing on the Sequel Systems
1. Single-Cell cDNA Synthesis
   - Can use any third-party single-cell isolation system that produces full-length cDNA containing cell barcodes and UMIs

2. Single-Cell cDNA Re-Amplification & ProNex Bead Purification
   - Follow Procedure & Checklist – Preparing Single-Cell Iso-Seq Libraries Using SMRTbell Express Template Prep Kit 2.0 (PN 101-892-000)
   - Intact (un-sheared) RT-PCR products initially generated using a third-party single-cell preparation system are re-amplified to increase the mass
   - Single-cell library amplification primers for a specific platform may be ordered from any oligo synthesis vendor
   - The amount of ProNex beads to use to purify the amplified cDNA depends on the distribution of transcripts produced by the single-cell cDNA preparation and the goal of the experiment.

3. SMRTbell Express TPK 2.0 Library Construction
   - Single-tube, addition-only reactions
   - Typical library yield ≥40%

4. Sequencing Preparation
   - Anneal sequencing primer, bind polymerase, perform ProNex bead complex cleanup
   - Generate up to ~3 Million full-length single-cell transcript reads per Sequel II SMRT Cell 8M

5. Data Analysis
   - De-multiplex barcodes within SMRT Link GUI or on the command line
   - Single-cell data analysis can be performed using third-party software such as SQANTI and cDNA_cupcake

- To target transcripts ≥2 kb or larger, use 86 µLof ProNex Beads
- To target transcripts <1 kb, use 95 µLof ProNex Beads

- DNA Damage Repair
- End Repair / A-Tailing
- Overhang Adapter Ligation
- ProNex Bead Purification

- Verifying Input Mass Requirements for Library Construction:
  - >80 ng cDNA for Sequel
  - >160 ng cDNA for Sequel II

- Data Analysis
  - De-multiplex barcodes within SMRT Link GUI or on the command line
  - Single-cell data analysis can be performed using third-party software such as SQANTI and cDNA_cupcake
# LIST OF REQUIRED EQUIPMENT AND MATERIALS FOR cDNA RE-AMPLIFICATION AND SMRTBELL LIBRARY CONSTRUCTION

<table>
<thead>
<tr>
<th>ITEM</th>
<th>VENDOR</th>
<th>PROTOCOL STEP(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TempAssure PCR 8-tube strips - 0.2 ml PCR 8-tube FLEX-FREE strip, attached flat caps are recommended; <strong>OR</strong></td>
<td>USA Scientific, Inc. – PN 1402-4708 (recommended)</td>
<td>cDNA Re-amplification; ProNex Bead Purification, SMRTbell Library Construction</td>
</tr>
<tr>
<td>0.2 ml 8-Tube PCR Strips without Caps 0.2 ml &amp; Domed PCR Tube 8-Cap Strips</td>
<td>Bio-Rad – PN TBS0201, TCS0801</td>
<td></td>
</tr>
<tr>
<td>HDPE 8 place Magnetic Separation Rack for 0.2 ml PCR Tubes (recommended); <strong>OR</strong></td>
<td>V&amp;P Scientific Inc. – PN VP772F4-1 (International and Domestic) or Fisher Scientific – PN NC0988547 (Domestic only)</td>
<td>ProNex Bead Purification</td>
</tr>
<tr>
<td>0.2 mL PCR Strip Magnetic Separator 8 or 12 Strip</td>
<td>Permagen Labware – PN MSR812</td>
<td></td>
</tr>
<tr>
<td>ProNex Beads (for size selection)</td>
<td>Promega - PNs: NG2001 - 10mL, NG2002 - 125mL, NG2003 - 500mL</td>
<td>ProNex Bead Purification</td>
</tr>
<tr>
<td>Elution Buffer (50 mL)</td>
<td>PacBio PN 101-633-500</td>
<td>ProNex Bead Purification</td>
</tr>
<tr>
<td>Ethanol</td>
<td>Any MLS</td>
<td>ProNex Bead Purification</td>
</tr>
<tr>
<td>Qubit dsDNA HS Assay Kit</td>
<td>Invitrogen</td>
<td>DNA quantitation QC</td>
</tr>
<tr>
<td>Qubit Fluorometer</td>
<td>Invitrogen</td>
<td>DNA quantitation QC</td>
</tr>
<tr>
<td>High Sensitivity DNA Kit</td>
<td>Agilent</td>
<td>DNA sizing QC</td>
</tr>
<tr>
<td>Bioanalyzer Instrument</td>
<td>Agilent</td>
<td>DNA sizing QC</td>
</tr>
<tr>
<td>8-channel pipettes for processing multiple samples (200 µL &amp; 20 µL)</td>
<td>Any MLS</td>
<td>cDNA Re-amplication; ProNex Bead Purification, SMRTbell Library Construction</td>
</tr>
<tr>
<td>Thermal cycler that is compatible with 8-tube strips (100 µL volume)</td>
<td>Any MLS</td>
<td>cDNA Re-amplication; SMRTbell Library Construction</td>
</tr>
<tr>
<td>SMRTbell Express Template Prep Kit 2.0</td>
<td>Pacific Biosciences</td>
<td>SMRTbell Library Construction</td>
</tr>
<tr>
<td>NEBNext High-Fidelity 2X PCR Master Mix (for additional PCR reactions)</td>
<td>NEB PN M0541S</td>
<td></td>
</tr>
<tr>
<td>Single-Cell cDNA library amplification primers for a specific platform may be ordered from any oligo synthesis company†</td>
<td>Single-cell vendor-specific</td>
<td>cDNA Re-amplification</td>
</tr>
</tbody>
</table>

**Note:** Commercial vendors may have different sequences for their PCR Primers. Please consult your vendor for the specific primer(s) sequence(s).

† Pacific Biosciences does not sell a kit for carrying out the Single-Cell RNA Sequencing method. Use of these methods may require rights to third-party owned intellectual property.

---

Example 1: 10X Chromium Single Cell 3' V2 and V3 PCR Primers  
Forward PCR primer (Primer 1):  
5'-CTACACGACGCTCTTCCGATCT-3'  
Reverse PCR primer (Primer 2):  
5'-AAGCAGTGTATGATGCTCAACCGAGGTAATGATGATGCTCAACGGAGGT-3'  
Example 2: Drop-Seq uses the same sequence for PCR Primer 1 and Primer 2:  
5'-AAGCAGTGTATGATGCTCAACCGAGGTAATGATGATGCTCAACGGAGGT-3'  
BEST PRACTICES RECOMMENDATIONS FOR PREPARING SINGLE-CELL ISO-SEQ SMRTBELL LIBRARIES

1. Always set your heat blocks or thermocyclers to the appropriate temperature for incubations **before** proceeding with the procedure.

2. ProNex beads:
   a. Equilibrate the ProNex beads at room temperature for 30 mins prior to use.
   b. It is critical to **accurately pipette** ProNex beads because small changes in volume can significantly alter the size distribution of your sample.

3. When performing ProNex bead purification steps, note that 80% ethanol is **hygroscopic** and should be prepared FRESH to achieve optimal results.

4. Using a multi-channel pipettor greatly enhances the ease of processing more than 1 sample.

5. Measure DNA concentration using a **Qubit fluorometer** and **Qubit dsDNA High Sensitivity (HS) Assay Kit** reagents as recommended by the manufacturer.
PREPARATION OF SINGLE-CELL cDNA PCR RE-AMPLIFICATION REACTION

PCR Reaction Preparation and Thermal Cycling Conditions

1. On ice, prepare Reaction Mix 1 by adding the following components in the order listed.

<table>
<thead>
<tr>
<th>Reaction Mix 1</th>
<th>Volume</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEBNext® High-Fidelity 2X PCR Master Mix</td>
<td>50 µL</td>
<td></td>
</tr>
<tr>
<td>12 µM PCR Primer 1</td>
<td>2 µL</td>
<td></td>
</tr>
<tr>
<td>12 µM PCR Primer 2</td>
<td>2 µL</td>
<td></td>
</tr>
<tr>
<td>Single-Cell cDNA</td>
<td>648 µL</td>
<td></td>
</tr>
<tr>
<td>H₂O</td>
<td>Up to 100 µL</td>
<td></td>
</tr>
<tr>
<td>Total Volume</td>
<td>100 µL</td>
<td></td>
</tr>
</tbody>
</table>

- Total single-cell (SC) cDNA input mass depends on the specific third-party SC platform and input # of cells used for the SC cDNA preparation.
- PacBio recommends using a maximum input of ~5000 cells per SC Iso-Seq library preparation.

2. Gently vortex by performing two 2-second pulses and then perform a quick spin to collect all liquid from the sides of the tube.

3. Place the tubes in a thermocycler and run the following program (lid temperature = 105°C):

<table>
<thead>
<tr>
<th>PCR Program</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>45 seconds at 98°C</td>
<td>1 cycle</td>
</tr>
<tr>
<td>10 seconds at 98°C</td>
<td></td>
</tr>
<tr>
<td>15 seconds at 62°C</td>
<td>6 cycles*</td>
</tr>
<tr>
<td>3 minutes at 72°C</td>
<td></td>
</tr>
<tr>
<td>5 minutes at 72°C</td>
<td>1 cycle</td>
</tr>
<tr>
<td>Hold at 4°C</td>
<td></td>
</tr>
</tbody>
</table>

- 6 cycles is a recommended starting point.
- If > 500 ng is routinely obtained, reduce the number of PCR cycles for future experiments.
- If < 200 ng is routinely obtained increase the number of PCR cycles for future experiments.

* Target yield is approximately 200 ng. The number of cycles will vary significantly across systems, number of cells, etc.
PURIFICATION OF RE-AMPLIFIED SINGLE-CELL cDNA PRODUCTS

The specific method chosen to purify the re-amplified single-cell cDNA depends on the goal of the experiment and the expected size distribution of transcripts.

- Use ProNex beads for purification of amplified cDNA products according to the table below:

<table>
<thead>
<tr>
<th>GOAL OF EXPERIMENT</th>
<th>PRONEX BEAD VOLUME</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Single-cell cDNA sample is composed of transcripts centered at ~2 kb or larger; or</td>
<td>86 µL</td>
</tr>
<tr>
<td>2) A reduction in shorter transcripts is desired.</td>
<td></td>
</tr>
<tr>
<td>1) Single-cell cDNA sample composed primarily of transcripts &lt;2 kb; or</td>
<td>95 µL</td>
</tr>
<tr>
<td>2) No reduction in shorter transcripts is desired.</td>
<td></td>
</tr>
</tbody>
</table>

- After purification, perform a sizing QC by running 1 µL of the purified cDNA products on a Bioanalyzer system using a High Sensitivity DNA kit.

- Examining the amplified cDNA on a Bioanalyzer prior to PacBio library construction is an excellent quality control step to ensure that the amplified cDNA material has the expected size distribution.
QUANTITATION OF RE-AMPLIFIED SINGLE-CELL cDNA PRODUCTS

Use a Qubit dsDNA HS assay kit to verify that you have the required mass of purified single-cell cDNA to proceed with SMRTbell library construction.

<table>
<thead>
<tr>
<th>INSTRUMENT</th>
<th>MINIMUM cDNA Sample Amount</th>
<th>RECOMMENDATION FOR SAMPLES WITH LOW YIELD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequel System</td>
<td>80 – 500 ng</td>
<td>If total mass is &lt;80 ng (&lt;1.75 ng/μL) → Go to Appendix 1</td>
</tr>
<tr>
<td>Sequel II and Ile Systems</td>
<td>160 – 500 ng</td>
<td>If total mass is &lt;160 ng (&lt;3.5 ng/μL) → Go to Appendix 1</td>
</tr>
</tbody>
</table>

- **Appendix 1: Recommendations for Additional cDNA Amplification by PCR for Samples with a Lower Yield:**
  - The Sequel System and Sequel II and Ile Systems require different amounts (ng) of cDNA for SMRTbell library construction. The Sequel System requires >80 ng of cDNA, while the Sequel II and Ile Systems require >160 ng of cDNA.
  - If there is not enough DNA to proceed with library construction, refer to Appendix 1 of the procedure which describes a workflow for enriching cDNA by performing additional PCR cycles.
  - **Note:** PCR over-amplification can result in sub-optimal data.
    - If >500 ng is routinely obtained after cDNA re-amplification, reduce the number of PCR cycles for future experiments.
SIZING QC ANALYSIS OF RE-AMPLIFIED SINGLE-CELL cDNA PRODUCTS

Examining the amplified cDNA on a Bioanalyzer system *prior* to proceeding with PacBio SMRTbell library construction is an excellent quality control step to ensure that the amplified cDNA material has the expected size distribution.

Bioanalyzer sizing QC analysis of a re-amplified single-cell cDNA sample. The re-amplified cDNA sample was purified using 0.95X ProNex beads (95 µL ProNex beads + 100 µL cDNA sample). The average size of the purified single-cell cDNA sample is 1.9 kb.
SMRTBELL EXPRESS TEMPLATE PREP KIT 2.0 REAGENT HANDLING RECOMMENDATIONS

- Several reagents in the kits are sensitive to temperature and vortexing
- PacBio highly recommends:
  - Never leaving reagents at room temperature
  - Working on ice at all times when preparing master mixes
  - Finger tapping followed by a quick-spin prior to use

### LIST OF TEMPERATURE-SENSITIVE REAGENTS INCLUDED IN SMRTBELL EXPRESS TPK 2.0.

<table>
<thead>
<tr>
<th>REAGENT</th>
<th>WHERE USED</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA Damage Repair Mix v2</td>
<td>DNA Damage Repair</td>
</tr>
<tr>
<td>End Prep Mix</td>
<td>End-Repair/A-tailing</td>
</tr>
<tr>
<td>Overhang Adapter v3</td>
<td>Ligation</td>
</tr>
<tr>
<td>Ligation Mix</td>
<td>Ligation</td>
</tr>
<tr>
<td>Ligation Additive</td>
<td>Ligation</td>
</tr>
<tr>
<td>Ligation Enhancer</td>
<td>Ligation</td>
</tr>
</tbody>
</table>
SIZING QC ANALYSIS OF FINAL SINGLE-CELL ISO-SEQ SMRTBELL LIBRARIES AFTER PRONEX BEAD PURIFICATION

Characterize the final Single-Cell Iso-Seq SMRTbell library on a Bioanalyzer system prior to proceeding with sequencing to ensure that the purified library template material has the expected size distribution.

Bioanalyzer sizing QC analysis of a purified Single-Cell Iso-Seq SMRTbell Express TPK 2.0 library sample. The final Single-Cell Iso-Seq library sample was purified using 1X ProNex beads. The average size of the purified SMRTbell library sample is 2.2 kb.
SINGLE-CELL ISO-SEQ SMRTBELL LIBRARY CONSTRUCTION YIELDS

- Overall, SMRTbell library yields in this Single-Cell Iso-Seq workflow are typically 50% starting from re-amplified single-cell cDNA input into the first enzymatic step (DNA Damage Repair).

- Sufficient amounts of SMRTbell template material can typically be generated using this procedure to run the following numbers of SMRT Cells (per Single-Cell Iso-Seq library preparation reaction):
  - **Sequel System:** ≥3 SMRT Cells 1M
  - **Sequel II and IIe Systems:** ≥1 SMRT Cells 8M
Single-Cell Iso-Seq Library Sequencing Workflow Details
SAMPLE SETUP RECOMMENDATIONS FOR SINGLE-CELL ISO-SEQ LIBRARIES – SEQUEL SYSTEM (CHEMISTRY 3.0)

- Follow SMRT Link Sample Setup instructions using the recommendations provided in the Quick Reference Card – Loading and Pre-Extension Time Recommendations for the Sequel System for sequencing bulk Iso-Seq samples.

* PacBio recommends Sequel Binding Kit 3.0 for all Iso-Seq Express workflows (Short, Standard, and Long)
SAMPLE SETUP RECOMMENDATIONS FOR SINGLE-CELL ISO-SEQ LIBRARIES – SEQUEL II AND Ile SYSTEMS (CHEMISTRY 2.0)

- Follow SMRT Link Sample Setup instructions using the recommendations provided in the Quick Reference Card – Loading and Pre-Extension Time Recommendations for the Sequel II and Ile Systems for sequencing bulk Iso-Seq samples.

- For SMRT Link v10.0 (or higher): Select ‘Iso-Seq Method’ from the Application field dropdown menu in the SMRT Link Sample Setup and SMRT Link Run Design user interface.

<table>
<thead>
<tr>
<th>Applications</th>
<th>Data Type</th>
<th>Library Prep Kit</th>
<th>Binding Kit</th>
<th>Sequencing Primer</th>
<th>Pol Binding Time (hr)</th>
<th>Complex Cleanup</th>
<th>Loading Concentration Range (pM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iso-Seq Method (standard samples)</td>
<td>CCS</td>
<td>Express Prep 2.0</td>
<td>Binding Kit 2.1</td>
<td>v4</td>
<td>1</td>
<td>1.2X ProNex Beads</td>
<td>40 - 80</td>
</tr>
<tr>
<td>Iso-Seq Method (focus on long transcripts)</td>
<td>CCS</td>
<td>Express Prep 2.0</td>
<td>Binding Kit 2.0</td>
<td>v4</td>
<td>1</td>
<td>1.2X ProNex Beads</td>
<td>50 - 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Applications</th>
<th>Pre-Extension Time (hr)</th>
<th>Movie Collection Time (hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iso-Seq Method (standard samples)</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>Iso-Seq Method (focus on long transcripts)</td>
<td>2</td>
<td>24</td>
</tr>
</tbody>
</table>

* PacBio recommends Sequel II Binding Kit 2.1 for standard bulk Iso-Seq and Single-Cell Iso-Seq samples. For bulk Iso-Seq and Single-Cell Iso-Seq samples with a focus on long transcripts, PacBio recommends Sequel II Binding Kit 2.0.
Single-Cell Iso-Seq Data Analysis Recommendations
SINGLE-CELL ISO-SEQ DATA ANALYSIS RECOMMENDATIONS AND GUIDELINES

Data analysis guidelines for Single-Cell Iso-Seq applications can be found on PacBio's GitHub website.

Iso Seq Single Cell Analysis: Recommended Analysis Guidelines
Elizabeth Tseng edited this page 3 days ago · 62 revisions

Last Updated: 01/15/2021

NOTE 1: The guidelines on this wiki are constantly evolving. Please check back often for updates! For any issues or bugs, please use issues with the title [SingleCell].

NOTE 2: You can use this wiki for both single cell (UMI+BC) and non-single cell data that only has UMIs.

1. Generate CCS Reads
2. Detect and Remove 5’ and 3’ Primers
3. Detect UMIs and Cell Barcodes
4. Remove PolyA Tail and Artificial Concatemers
5. Cluster Reads by Unique Founder Molecules
6. Align to Genome
7. Collapse into Unique Transcripts
8. Compare Against Annotation
9. Filter Artifacts
10. Process into CSV Report and UMI/BC Error Correction

STRUCTURAL DIFFERENCES BETWEEN BULK ISO-SEQ VS. SINGLE-CELL ISO-SEQ TRANSCRIPT DATA

**Bulk Iso-Seq Transcripts**

1. Poly(A+) mRNA
2. Full-Length 1st-Strand cDNA Synthesis
3. Full-Length cDNA Amplification
4. SMRTbell TPK 2.0 Library Construction
5. SMRT Sequencing

**Single-Cell Iso-Seq (scIso-Seq) Transcripts**

Any single-cell platform that generates full-length cDNA can be used for sequencing with PacBio

1. Full-Length cDNA Re-Amplification
2. SMRTbell TPK 2.0 Library Construction
3. SMRT Sequencing

Transcript structure depends on single-cell platform used.

- **Bulk Iso-Seq**:
  - Use Barcoded PCR Primers if Multiplexing Bulk Iso-Seq Samples
  - Pool Barcoded cDNA if Multiplexing

- **Single-Cell Iso-Seq**:
  - Use PCR Primers Specific to single-cell Isolation System

**Primer Names**

- 5’primer
- transcript
- (A)n
- 3’primer

**Library**

- SMRTbell TPK 2.0

**Sequencing**

- SMRT Sequencing
SINGLE-CELL ISO-SEQ DATA ANALYSIS WORKFLOW SUMMARY OVERVIEW

Single-Cell Iso-Seq Bioinformatics Tutorial on Github assumes that the UMIs and cell barcodes (BCs) are on the 3' end (between the polyA tail and 3' primer). The UMIs and BCs can be of any length.

1. Generate CCS Reads
   - 5' primer \(\rightarrow\) transcript \(\rightarrow\) \((A)n\) \(\rightarrow\) UMI \(\rightarrow\) Cell BC \(\rightarrow\) 3' primer

2. Detect and Remove 5' and 3' Primers
   - transcript \(\rightarrow\) \((A)n\) \(\rightarrow\) UMI \(\rightarrow\) Cell BC

3. Detect UMIs and Cell Barcodes
   - transcript \(\rightarrow\) \((A)n\)

4. Remove polyA Tail and Artificial Concatemers
   - transcript

5. Cluster Reads by Unique Founder Molecules
   - Performs PCR de-duplication via clustering by UMI and cell barcodes (if available) using `dedup`. After deduplication, `dedup` generates one consensus sequence per founder molecule, using a QV guided consensus approach

The following Single-Cell Iso-Seq analysis steps below are **optional**.

1. **Align to Genome**
   - Use `minimap2` (included in SMRT Analysis), desalt, GMAP, STAR, or BLAT etc. See aligner tutorial on PacBio GitHub website for additional recommendations.

2. **Collapse into Unique Transcripts**
   - Use `Cupcake` to collapse mapped FLNC reads into unique transcripts. Read the `Cupcake` wiki for more details on each script.

3. **Compare Against Annotation**
   - Use `SQANTI3` to annotate each unique transcript against an annotation. To run SQANTI3, you must provide a GTF annotation (such as GENCODE) and a reference genome FASTA.

4. **Filter Artifacts**
   - **OPTIONAL**: Can use `SQANTI3` to remove cDNA library construction artifacts (e.g., intra-priming artifacts, RT switching artifacts, TSO artifacts, etc.)

5. **Process into CSV Report and UMI/BC Error Correction**
   - Can use `cDNA_Cupcake/single-cell` scripts to generate a collated CSV file that links each mapped FLNC read to its classified genes and transcripts based on `SQANTI3`'s output.

## TYPES OF ISO-SEQ cDNA LIBRARY ARTIFACTS

### Expected Full-Length Iso-Seq Transcript Structure

<table>
<thead>
<tr>
<th>5’ Primer</th>
<th>ATGGG</th>
<th>Transcript</th>
<th>(AAAA)n</th>
<th>3’ Primer</th>
</tr>
</thead>
</table>

### RT Priming Artifact

<table>
<thead>
<tr>
<th>3’ Primer</th>
<th>(TTTT)n</th>
<th>Transcript</th>
<th>(AAAA)n</th>
<th>3’ Primer</th>
</tr>
</thead>
</table>

3’ – 3’ primer artifact can occur if TSO was not added

### TSO Priming Artifact

<table>
<thead>
<tr>
<th>5’ Primer</th>
<th>ATGGG</th>
<th>Transcript</th>
<th>CCCAT</th>
<th>5’ Primer</th>
</tr>
</thead>
</table>

5’ – 5’ primer artifact can occur if TSO acts as a primer on the RNA

### RT Switching Artifact (Genomic View)

- **EXON 1**
- **EXON 2**

This part of an exon goes missing and when re-mapped to the genome looks like a new intron

### Intra-priming Artifact (Genomic View)

- **EXON 1**
- **EXON 2**

Priming off genomic ‘A’s results in false 3’ end
RT primer (3’p – 3’p) artifacts can occur if TSO was not added

Template switching occurs followed by 2nd-strand cDNA synthesis and amplification

No template switching occurs since TSO is absent. RT Primer primes 2nd-strand cDNA synthesis followed by amplification with Reverse (3’) Primer on both ends.
TYPES OF ISO-SEQ cDNA LIBRARY ARTIFACTS (CONT.)

TSO primer (5’p – 5’p) artifacts can occur if the TSO is acting as a primer on the RNA

Template switching occurs followed by
2nd-strand cDNA synthesis and amplification

Reverse (3’)
PCR Primer

Desired Product

RT Primer

TSO Primer Artifact

TSO primer (5’p – 5’p) artifacts can occur if the TSO is acting as a primer on the RNA.
Example LIMA report output indicating that the majority of reads that did not pass QC show a 5'p – 5'p artifact. Such types of 5'p – 5'p artifacts typically result when the TSO is acting as a primer on the input RNA template during first-strand cDNA synthesis.
SQANTI: A TOOL FOR CLASSIFYING FULL-LENGTH TRANSCRIPTS

LONG READ DEFINED TRANSCRIPTS
FASTA / FASTQ / GTF

REFERENCE GENOME
FASTA

REFERENCE TRANSCRIPTOME
GTF
Gene A: 
Isoform #1
Isoform #2

OPTIONAL INPUT
• RNA-Seq
• CAGE peak data
• polyA motifs
• tappAS-like functional annotation...

Classification file
<table>
<thead>
<tr>
<th>ID</th>
<th>Gene</th>
<th>Transcript</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB.1.1</td>
<td>Ctnnd1</td>
<td>ENSMUST000000067232</td>
<td>FSM</td>
</tr>
<tr>
<td>PB.1.2</td>
<td>Ctnnd1</td>
<td>novel</td>
<td>NIC</td>
</tr>
<tr>
<td>PB.2.1</td>
<td>Novel</td>
<td>novel</td>
<td>Intergenic</td>
</tr>
</tbody>
</table>

Junction file
<table>
<thead>
<tr>
<th>Junction</th>
<th>Isoform</th>
<th>Splice site</th>
<th>Known</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junction1</td>
<td>PB.1.2</td>
<td>GT-AG</td>
<td>True</td>
</tr>
<tr>
<td>Junction2</td>
<td>PB.1.2</td>
<td>GC-AG</td>
<td>True</td>
</tr>
<tr>
<td>Junction3</td>
<td>PB.1.2</td>
<td>GT-AF</td>
<td>False</td>
</tr>
</tbody>
</table>

PDF Report

Complementary files
• Genome-corrected transcriptome
• ORF prediction (FASTA)
• CDS-annotated GTF
• tappAS-compatible GFF3

https://github.com/ConesaLab/SQANTI3

**SQANTI CLASSIFICATION OF ISO-SEQ TRANSCRIPTS**

- **Full Splice Match (FSM)**
  Matches all splice junctions (SJs) perfectly

- **Incomplete Splice Match (ISM)**
  Matches reference the SJs partially

- **Novel In Catalog (NIC)**
  Novel isoform with a new combination of known SJs

- **Novel Not in Catalog (NNC)**
  Novel isoform with at least one new SJ

- **Genic Intron**
  Within an intron

- **Genic Genomic**
  Overlaps with introns and exons

---

USING SQANTI AS AN ISO-SEQ DATA QUALITY CONTROL TOOL

What the SQANTI tool is intended for:
- Classification of isoforms against known transcripts
- Classification of splice junctions
- Identification of potential cDNA library artifacts
- Aggregation of supporting evidence for full-length transcripts
- Supporting evidence for functionality (ORFs, Nonsense-mediated mRNA decay (NMD))

What the SQANTI tool is **NOT** intended for:
- Fixing problems with other parts of the analysis pipeline (e.g., aligner mistakes, low quality sequences, multi-mapping issues)
Single-Cell Iso-Seq Library Example Sequencing Performance Data
COMPARISON OF cDNA TRANSCRIPT LENGTH DISTRIBUTIONS FOR SINGLE-CELL ISO-SEQ LIBRARY PREPARATIONS VS. BULK ISO-SEQ LIBRARY PREPARATIONS

Single-Cell cDNA samples prepared using third-party single-cell isolation platforms tend to produce shorter transcript lengths than bulk cDNA preparations*

- Bulk cDNA Prep Alzheimer brain: 80 – 14,288 bp (Mean Length: 3.3 kb)
- Single-Cell cDNA Prep (Platform 1) Human Brain Organoid: 80 – 8,607 bp (Mean Length: 1.7 kb)
- Single-Cell cDNA Prep (Platform 2) Human Cell Line: 80 – 5,834 bp (Mean Length: 1.4 kb)

* Read lengths, reads/data per SMRT Cell and other sequencing performance results vary based on sample quality/type and insert size.
COMPARISON OF ISO-FORM DISTRIBUTIONS ACROSS STRUCTURAL CATEGORIES FOR HUMAN SINGLE-CELL ISO-SEQ SAMPLES

Transcript isoform distribution is highly sample-dependent

**Single-Cell, Human Organoid**

**Single-Cell, Human Cell Line**
FULL-LENGTH SINGLE-CELL ISO-SEQ METHOD CAN REVEAL CELL-TYPE SPECIFIC ISOFORM EXPRESSION

Example: Multiple tropoelastin isoforms are found in astrocytes but not in other central nervous system (CNS) tissue cell types

Tropoelastin gene is expressed only in astrocytes but not other human organoid cell types and shows exon skipping events and usage of alternative start/end sites.
FULL-LENGTH SINGLE-CELL ISO-SEQ METHOD CAN REVEAL SPECIES-SPECIFIC ISOFORM EXPRESSION

Example: Alternative transcription start site (TSS) usage for ZNF331 gene differs in chimp vs. human cerebral organoids

Human cerebral organoids show a preference for utilization of an upstream transcription start site in the ZNF331 gene, which may be a contributing factor to the higher expression levels of ZNF331 in humans vs. chimps.
Technical Documentation & Applications Support Resources
**BEST PRACTICES:** SINGLE-CELL RNA SEQUENCING (scISO-SEQ ANALYSIS)

**Template Preparation with SMRTbell Express Template Prep Kit 2.0**
- Enrich for single-cell cDNA using a single-cell sorting platform that generates full-length cDNA*
  - Template switch oligo (TSO)-based cDNA synthesis methods are recommended
  - The final single-cell cDNA product consists of 5’ primer, transcript, poly-A tail, unique molecular index (UMI), cell barcode and 3’ primer
  - To generate matching short-read data, save 5% of the material
  - Additional PCR cycles can be added if necessary
- Start library preparation with at least 160 ng of input cDNA (post-single-cell platform PCR reaction) for 1-2 SMRT Cell 8M
  - More starting material will be required for sequencing multiple SMRT Cells 8M
- Prepare libraries with the SMRTbell Express Template Prep Kit 2.0 in one day

**Sequence on the Sequel, Sequel II or Sequel Ile System**
- Use the Sequel II or Ile Systems to generate ~3 million full-length reads from one SMRT Cell 8M to obtain ~1,000 unique molecules for 3,000 single cells**
  - Use 24 hr movies with 2 hrs pre-extension time
  - For human samples, run up to 240 SMRT Cell 8M/year at a cost of ~$1,300/SMRT Cell 8M, excluding single-cell enrichment cost†

**Data Analysis Solutions with the PacBio Analytical Portfolio**
- Analyze HiFi reads which allow accurate single-cell barcode and UMI identification
- Use the single-cell Iso-Seq analysis tools on GitHub to output high-quality, full-length transcript FASTA sequences per UMI, with no assembly required, to characterize transcript variants for each cell

---

* Number of usable reads, containing the UMI and cell barcode, vary by single-cell platform. Any platform that generates full-length cDNA is compatible with the single-cell RNA sequencing workflow.

** Read lengths, reads/data per SMRT Cell type and other sequencing performance results vary based on single-cell platform, sample quality/type and insert size.

† Prices, listed in USD, are approximate and may vary by region. Pricing includes library and sequencing reagents run on a Sequel II or Ile System and does not include instrument amortization or other reagents.
Sample Preparation Literature

- Procedure & Checklist – Preparing Single-Cell Iso-Seq Libraries Using SMRTbell Express Template Prep Kit 2.0 (PN 101-892-000)
- Quick Reference Card – Loading and Pre-extension Recommendations for the Sequel System (PN 101-461-600)
- Quick Reference Card – Loading and Pre-extension Recommendations for the Sequel II System (PN 101-769-100)
- Overview – Sequel Systems Application Options and Sequencing Recommendations (PN 101-851-300)
- Application Consumable Bundles Purchasing Guide (PN PG100-051320)
- Technical Overview: Single-Cell Iso-Seq Express Library Preparation Using SMRTbell Express Template Prep Kit 2.0 (PN 101-925-400)

Posters

- PacBio AGBT 2020 Poster: A Complete Solution for Full-Length Transcript Sequencing Using the PacBio Sequel II System
- PacBio ENCODE 2019 Poster: Single Cell Isoform Sequencing (scIso-Seq) Identifies Novel Full-length mRNAs and Cell Type-specific Expression
TECHNICAL DOCUMENTATION AND APPLICATIONS SUPPORT RESOURCES FOR SINGLE-CELL ISO-SEQ LIBRARY PREPARATION, SEQUENCING & DATA ANALYSIS (CONT.)

Videos & Webinars

- PacBio ASHG 2020 CoLab: PacBio HiFi reads for comprehensive characterization of genomes and single-cell isoform expression
- PacBio ASHG 2020 Workshop: Single-cell isoform analysis of the nervous system
- LabRoots 2020 Presentation: Single cell gene expression: new insights through the lens of full-length mRNA isoform resolution
- PacBio AGBT 2019 Presentation: Single cell isoform sequencing (sclIso-Seq) identifies novel full-length mRNAs

Publications

Data Analysis Resources

- Single-Cell Iso Seq Analysis Tutorial on GitHub: Recommended scIso-Seq Analysis Guidelines
  - **NOTE:** The Single-Cell Iso-Seq data analysis guidelines presented in this wiki document are **not** officially supported by PacBio.