



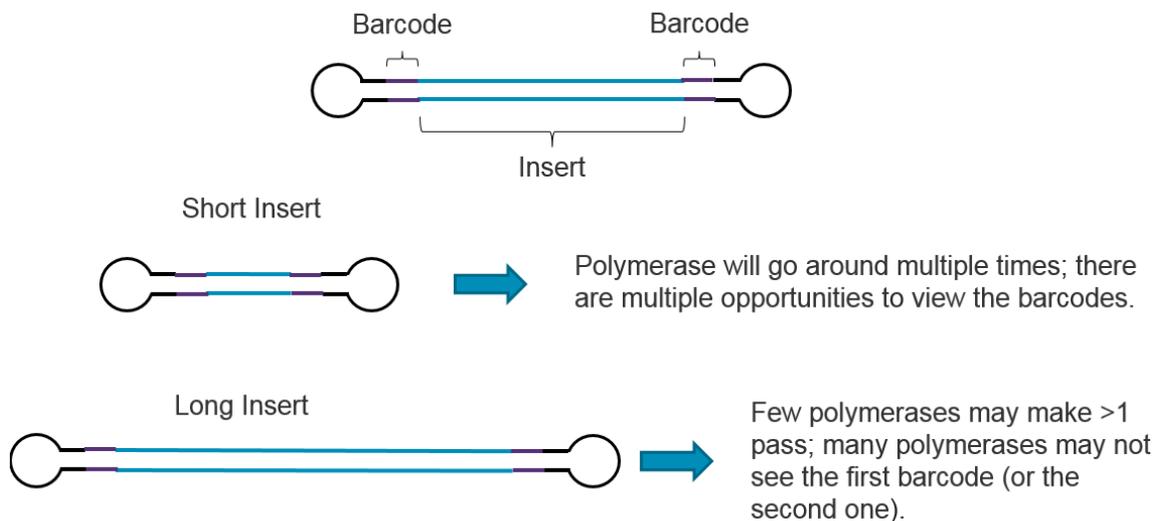
Introduction

This document is for users with Sequel[®] Systems using SMRT Link v5.0.0 or v5.0.1. This document covers:

- Barcoding designs, strategies and modes for preparing barcoded samples and data analysis for multiplexed sequencing for the following applications: Amplicons (Minor variants, LAA and CCS) and microbial assembly.
- Identifying barcodes using SMRT Link.
- Performing downstream data analysis using SMRT Link.

This document does **not** include information about barcoding designs and analysis for the Iso-Seq[®] application. For details about barcoding Iso-Seq samples, see <http://www.pacb.com/documentation/unsupported-protocol-barcoding-samples-for-isoform-sequencing-iso-seq%C2%99analysis/>.

Barcode Background

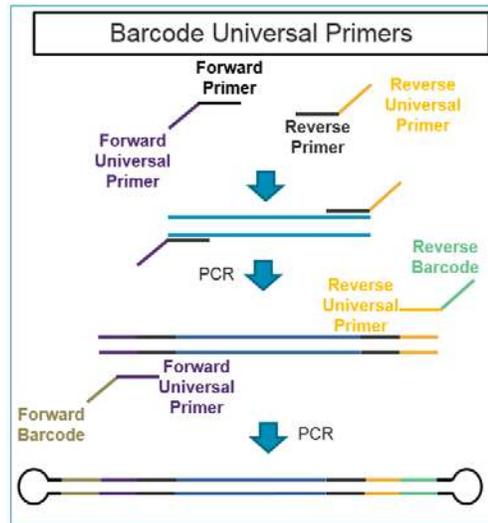
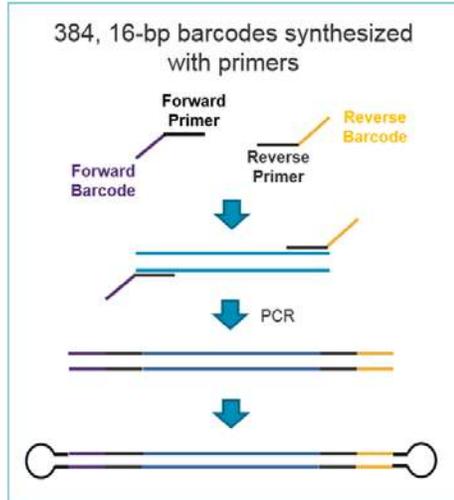


Multiplexed Analysis in SMRT Analysis - Matching Sample Preparation to Secondary Analysis

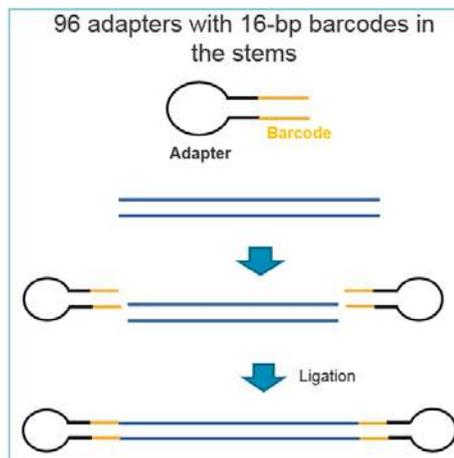
Barcoded analysis in SMRT Link requires users to properly match input barcode analysis sets to the expected **barcodes** and **barcode sequence orientation** of the prepared library so as to correctly score and demultiplex subreadsets. This document helps to identify which barcodes to use or upload to SMRT Link, for a given sample preparation.

Barcodes can be appended to a sample using three methods:

Barcoding during amplifications, 1 or 2 steps:



Barcoding after amplification or fragmentation:

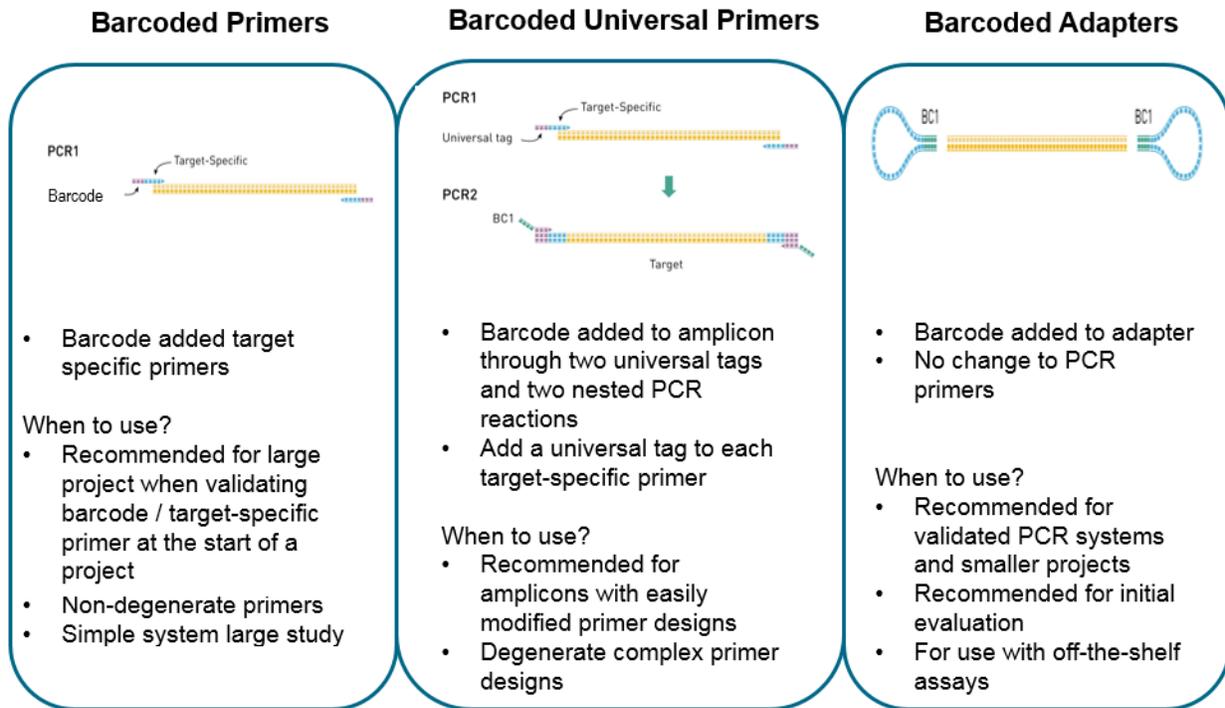


Barcoding Strategies

Three general purpose strategies are available for sequencing of targeted amplicons:

- Barcoded Primers
- Barcoded Universal Primers
- Barcoded Adapters

Additionally, the barcoded adapters strategy can be used for sheared libraries when multiplexing for microbial assembly projects.



Which Barcoding Strategies to Use?

Application	Minor Variant Analysis	CCS2	LAA2	Iso-Seq	Multiplexed HGAP
Library size	<4 kb amplicon	<4 kb amplicon	Amplicon(s) dependent	<10 kb cDNA	Sheared gDNA
Barcoded primers	Yes	Yes	Yes	Yes ¹	N/A
Barcoded Universal Primers	Yes	Yes	Yes	N/A	N/A
Barcoded Adapters	No	Yes ²	Yes	N/A	16
Multiplexing level available	≤384*384 samples	≤384*384 samples	≤384 samples	≤6 samples	2-16 samples

1 - Using 3' barcoded primer as per barcoded Iso-Seq[®] protocol.

2 - Can see potentially up to 5% barcode crossover.

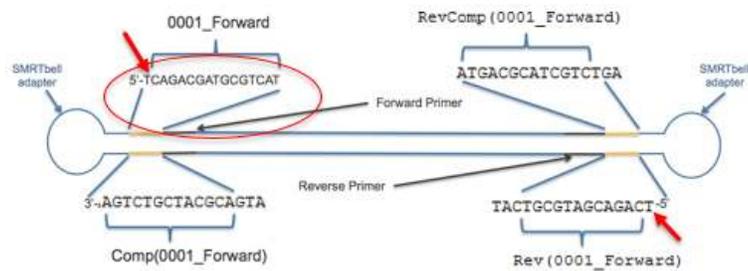
Barcode Scoring Modes and Barcode Sequence Files in SMRT Link

The above three barcoding strategies generate SMRTbell® templates with the barcode sequence directly **adjacent** to the SMRTbell adapters on **both** ends of the insert. This is the default configuration for most SMRT Analysis applications.

Subread movies in BAM format from either the PacBio RS II or Sequel System using these strategies can be analyzed using the **Barcoding** application in SMRT Link when coupled with the correct **barcode sequence file** and **scoring mode**.

Symmetric Mode

Barcode sequences are the **same** on both sides of the insert and in the **same orientation** with respect to the SMRTbell adapter. This is most commonly the case when using barcoded SMRTbell adapters and for target enrichment (non-hairpin) adapters. This is also the default scoring mode in SMRT Link.



SMRT Link v5.0.1 comes with a pre-installed barcode sequence file named `RSII_barcodes_revcomp` for use with PacBio barcoded adapter kits.

Custom Symmetric Barcode FASTA

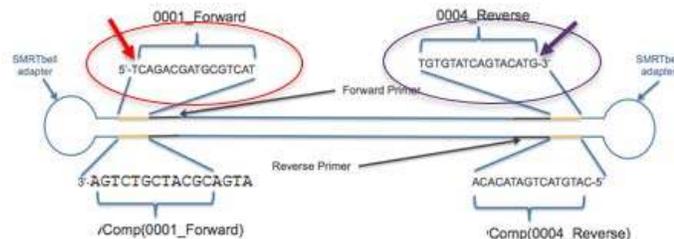
If you are using custom symmetric barcodes, you need to generate a FASTA file of barcode sequences. Make sure that the barcode sequences are listed in 5' - 3' with respect to the indicated strand. Note that the diagram is shown in the 5' - 3' direction.

```
1 >BC1
2 TCAGACGATGCGTCAT
3 >BC2
4 ACACACAGACTGTGAG
5 >BC3
6 TAIGACTAGTGTACTA
7 >BC4
8 TACGATCGTAGCTGCT
9 >BC5
10 CAGATCTCTCTGATGT
11 >BC6
12 ACGACAGTCAGAGTAT
```

Asymmetric Mode

Barcode sequences are **different** on the forward and reverse ends of the insert. Asymmetric mode is most commonly used when appending barcodes during a single round of PCR with barcoded primers. Pacific Biosciences recommends using this mode **only** for cases when both ends of the insert are expected to be sequenced for most molecules in the SMRT Cell (typically for amplicons <4 kb).

When running the **Barcoding** applications in SMRT Link, set the **Score Mode** in the **Advanced Analysis Parameters** dialog to `asymmetric`.



Custom Asymmetric Barcode FASTA

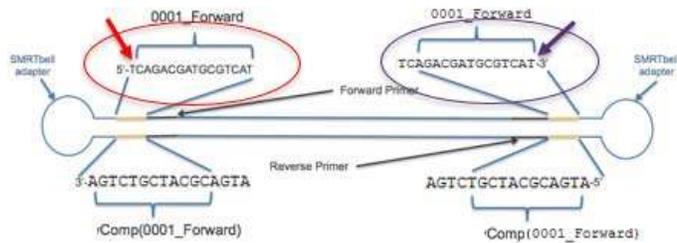
If you are using custom asymmetric barcodes, you need to generate a FASTA file of barcodes where each **sequential pair** of records is associated with **one** sample. Make sure that the barcode sequences are listed in 5' - 3' with respect to the indicated strand. Note that the diagram is shown in the 5' - 3' direction.

```
1 >FwdBC1
2 TCAGACGATGCGTCAT
3 >RevBC4
4 TGTGTATCAGTACATG
5 >FwdBC2
6 TATGACTAGTGTACTA
7 >RevBC5
8 TACGATCGTAGCTGCT
9 >FwdBC3
10 CAGATCTCTCTGATGT
11 >RevBC6
12 ACGACAGTCAGAGTAT
```

Barcode Universal Primers (BUP) and Barcoded Custom Primers

Barcode sequences are the **same** on both sides of the insert, but **opposite orientation** with respect to which end is abutting the SMRTbell adapter.

Note: This design should be scored using the `tailed` score mode along with either the `Sequel_RSII_96_barcode_v1` or `Sequel_RSII_384_barcode_v1` barcode sequence file, depending on the level of multiplex.



SMRT Link v5.0.1 comes with a pre-installed barcode sequence file named `pacbio_barcodes_universal` for use with BUP.

Note: Users who are using BUP reagents purchased **before** the v5.0.1 release should use the barcode sequence file `RSII_96_barcodes` with the Score Mode set to **tailed**.

Custom BUP Barcode FASTA

If you are using custom symmetric barcodes, you need to generate a FASTA file of barcode sequences. Make sure that the barcode sequences are listed in 5' - 3' with respect to the indicated strand. Note that the diagram is shown in the 5' - 3' direction.

```

1 >BC1
2 TCAGACGATGCGTCAT
3 >BC2
4 ACACACAGACTGTGAG
5 >BC3
6 TATGACTAGTGTACTA
7 >BC4
8 TACGATCCTAGCTGCT
9 >BC5
10 CAGATCTCTCTGATGT
11 >BC6
12 ACGACAGTCAGAGTAT

```

The following table includes the names of FASTA files used for analysis:

Fasta File	Sample Preparation			Score Mode		
	Barcoded Adapters	Barcoded Universal Primers	Custom Barcoded Primers	Symmetric	Asymmetric	Tailed
<code>RSII_96_barcodes</code>			✓	✓		✓
<code>RSII_96_barcodes_revcomp*</code>	✓			✓		
<code>RSII_384_barcodes</code>			✓	✓		✓
<code>RSII_384_barcodes_universal</code>		✓			✓	
<code>Sequel_RSII_16_barcodes_v1</code>			✓	✓		✓
<code>Sequel_RSII_96_barcodes_v1</code>			✓	✓		✓
<code>Sequel_RSII_384_barcodes_v1</code>			✓	✓		✓

✓ = Allowed combination for this file
✓ = Only Allowed Combination for this file
* = Use with **all** barcoded adapters

The following table describes the recommended Barcode file to use with different types of sample preparation:

Sample Preparation	Recommended Barcode File	Recommended Score Mode
Barcode Adapters	RSII_96_barcode_revcomp	Symmetric
Barcode Universal Primers	RSII_384_barcode	Tailed
Custom Barcoded Primers	Custom	Depends on Sample Preparation

The following table describes the recommended combinations of

- Platform and SMRT Analysis versions,
- Primers/Adapters, and
- Mode and Barcode file to use.

Platform/SMRT Analysis Version	BUP or Barcoded Primers (Same barcode, forward and reverse have different orientation)		BUP or Barcoded Primers (Different barcode on either side)		Barcoded Adapters (Same barcode and orientation forward and reverse)	
	Mode	Barcode File	Mode	Barcode File	Mode	Barcode File
RS II/v2.3.0	Symmetric	PacBio/Custom	Paired	Custom	Symmetric	PacBio/Custom
RS II or Sequel/v4.0.0	Asymmetric	Contains "universal"	Asymmetric	Custom	Symmetric	Contains "revcomp"
RS II or Sequel/v5.0.0	Tailed	PacBio/Custom	Asymmetric	Custom	Symmetric	Contains "revcomp"

Identifying Barcodes Using SMRT Link

The first step is to **import** the FASTA file with barcodes into SMRT Link.

Importing Barcode FASTA Files Into SMRT Link

You can import FASTA-format (.fa or .fasta) files containing barcodes for use in SMRT Link.

1. On the SMRT Link Home Page, select **Data Management**.
2. Click **View or Import Sequence Data**.
3. Click **Import**, then select **Barcodes (FASTA)**.



4. Navigate to the appropriate file and click **Import**. The FASTA file with barcodes is imported to the SMRT Link server and becomes available in SMRT Link. The barcode sequence file is labeled with the name of the barcode FASTA file.

After importing the barcodes, run the **Barcoding** application to **identify** the barcodes.

-
- The **Barcoding** application creates a new, **single** Data Set, which contains a reference to the barcode file used. The new Data Set is named `<your previous dataset name>(barcoded)`, and is the input for downstream analyses.
 - For **each** read, the new Data Set contains a **barcode** and a **barcode quality tag**. For example: `«bc:B:S,81,81»` and `«bq:i:38»` tags for symmetric barcode index 81 and a barcode quality score of 38.
 - The best barcode score for every ZMW is recorded in the `subreads.bam` file. Any identified barcode sequences are moved into the `scraps.bam` file.
 - For full information on BAM coding for barcodes, please refer to our BAM specifications:
<http://pacbiofileformats.readthedocs.io/en/5.0/BAM.html>.
 - **Note:** No barcode score filtering is performed at this step.

bam2bam

The Barcoding application calls the `bam2bam` utility:

Required inputs:

- `Subreads.bam`
- `Scraps.bam`
- `Barcodes.fasta` file
- Desired name for the output Data Set

Outputs:

- Labeled `subreads.bam`
- Labeled `scraps.bam`
- Labeled `subreadset.xml` file – this is the input for downstream applications.

Scoring mode:

- `symmetric`, `asymmetric`, and `tailed`.

Barcode Identification Results

- FASTA headers for barcodes are retained in the results.
- Barcode indices are still used for filtering and subsetting.

Barcodes			
Barcode Name	Polymerase Reads	Subreads	Bases
lbc1--lbc1	235	653	1,311,241
lbc10--lbc10	2,047	7,674	17,854,484
lbc11--lbc11	1,561	5,770	13,313,830
lbc12--lbc12	1,864	6,850	16,605,765
lbc13--lbc13	1,694	6,513	14,471,084
lbc14--lbc14	2,758	10,241	24,512,864
lbc15--lbc15	1,874	7,286	16,906,300
lbc16--lbc16	980	3,613	8,372,922
lbc17--lbc17	2,855	11,229	25,599,541
lbc18--lbc18	2,152	7,929	17,903,705
lbc19--lbc19	1,932	7,426	17,131,967
lbc2--lbc2	228	582	1,224,549
lbc20--lbc20	930	3,428	7,778,308
lbc21--lbc21	2,215	8,269	20,496,344
lbc22--lbc22	3,552	14,523	31,144,017
lbc23--lbc23	2,067	7,442	17,254,325
lbc24--lbc24	2,831	10,902	25,958,479
lbc25--lbc25	2,141	8,644	18,474,041
lbc26--lbc26	1,742	6,649	15,477,209

Working with Barcodes Downstream

The following SMRT Link analysis applications support (barcoded) Data Sets as input, and generate output data per barcode:

- Circular Consensus Sequences (CCS 2)
- Long Amplicon Analysis (LAA 2)
- Convert BAM to FASTX

The following analysis applications do **not** have any intrinsic barcode recognition. By default, they use **every** read available in the Data Set used as input, regardless of any barcode labels. To perform per-barcode analyses using these applications, first apply the barcode filter to the input Data Sets. (See the example on page 11.)

- Assembly HGAP4
- Resequencing
- Minor Variants Analysis [Beta]

The following analysis applications support 1-step barcode identification and sample analysis:

(Note: The barcode identification step does **not** separate out the data by barcode when used in a 1-step program; the data are all still together, just with a tag identifying the barcode.)

- CCS with Barcoding
- LAA with Barcoding

- Minor Variants Analysis with Barcoding [Beta]
- Iso-Seq[®], Iso-Seq[®] Classify Only, Iso-Seq[®] with Mapping.
Note: Barcoding for Iso-Seq samples is **not** the same as the methods described earlier in this document. For details, see <http://www.pacb.com/documentation/unsupported-protocol-barcoding-samples-for-isoform-sequencing-iso-seq%C2%99analysis/>.

Performing 1-Step Barcode Identification and Sample Analysis

1. Specify the barcode sequence file.
2. Specify the Scoring Mode: *symmetric*, *asymmetric*, or *tailed*.

3. Click **Advanced Analysis Parameters**.
4. Ensure that **Disable Barcode Filtering** is set to **OFF**.
5. Specify the minimum barcode alignment score. (Make sure to set it to 26 or larger.)

Note: Records have **not yet** been filtered based on barcode alignment score; this is all done during the actual analysis.

6. Click **OK**, then click **Start** to start the analysis.

Example: Performing HGAP Analysis Using a Barcoded Sub-read Set

1. In SMRT Link, click **SMRT Analysis**, then click **+ Create New Analysis**.
2. Enter a **name** for the analysis.
3. Specify whether to use **BAM-format** data generated on the Sequel System, or **PacBio RS II** data.
4. In the **Data Sets** box, select a barcoded subreadsets that includes the label (*barcoded*) in its name.
5. Select **Assembly (HGAP 4)** from the Applications droplist.
6. Specify the appropriate **Genome Length**.
7. Click **Advanced Analysis Parameters**.

8. Add comma-separated filters to the input data set in the **Filters to add to the dataset** field:

Example A (symmetric, tailed):
rq >= 0.7, bcf = 0, bcr = 0, bq >= 45

Example B (asymmetric):
rq >= 0.7, bcf = 0, bcr = 1, bq >= 45

- **rq**: Read Quality filter. Default setting is usually sufficient.
- **bcf**: Barcode Forward. The 0-indexed position of the forward barcode in the BarcodeSet used for barcoding.
- **bcr**: Barcode Reverse. The 0-indexed position of the reverse barcode in the BarcodeSet used for barcoding.
- **bq**: Barcode Quality Score. Software alignment score of the barcode on integer scale of 0-100. Recommended minimum: 45.

9. Change any other parameters and click **Start**.

10. Repeat steps 1-9 for **each barcode** in the Data Set.

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2017, Pacific Biosciences of California, Inc. All rights reserved. Information in this document is subject to change without notice. Pacific Biosciences assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of Pacific Biosciences products and/or third party products. Please refer to the applicable Pacific Biosciences Terms and Conditions of Sale and the applicable license terms at <http://www.pacb.com/legal-and-trademarks/product-license-and-use-restrictions/>.

Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science, Inc. NGS-go and NGSengine are trademarks of GenDx. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies. All other trademarks are the sole property of their respective owners.

P/N 101-307-200-01