Introduction

This document applies to Sequel® and Sequel II Systems using SMRT® Link v9.0.

Note: For information on SMRT Link v8.0, see the document **SMRT Analysis Barcoding Overview v8.0**, available here.

SMRT Link v9.0 includes a streamlined end-to-end barcoding workflow for automated analysis of multiplexed samples. It enables users to define barcodes and their associated sample names for an experiment at Run Design.

This workflow provides extensive QC metrics for evaluation of barcoding performance.

The SMRT Link graphical user interface supports up to 10,000 barcodes per sample. Support for more than 10,000 samples is available using the command-line. See the document **SMRT® Tools Reference Guide (v9.0)**, available here.

This document covers:

- Barcoding designs, strategies and modes for preparing barcoded samples and data analysis for **all** the applications available with SMRT Link v9.0.
- Identifying barcodes using SMRT Link.
- Finding demultiplexed `subread.bam` files.

This document does **not** include information about barcoding designs and analysis for the Iso-Seq® and No-Amp application. For details about barcoding samples for Iso-Seq analysis, see the document **SMRT Link User Guide (v9.0)**, available here.

PacBio Barcoding Background
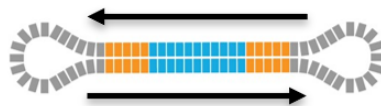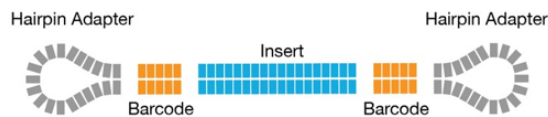
Multiple passes around a SMRTbell® molecule allows barcode detection.

Barcoding Strategies

There are several options for barcoding samples for SMRT sequencing:

1. **Barcoded Primers**
   – Target-specific primers are tailed with 16-bp barcodes. Use this method if your primers are already fully validated.

2. **Barcoded Universal Primers**
   – This method requires two rounds of PCR. The first round uses target-specific primers tailed with universal forward and reverse sequences.
   – For the second PCR round, round one PCR products are barcoded using F/R universal primers tailed with 16-bp barcodes.
   – For **symmetric** barcoding, PacBio recommends using our commercially-available 96 Barcoded Universal Primers kit. For **asymmetric** barcoding, we recommend using the barcoded M13 Primer approach.
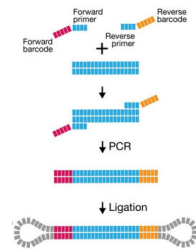
3. **Barcoded Overhang Adapters**
   – PacBio barcodes are added to the sample through ligation of barcoded overhang SMRTbell® adapters.
   – Recommended for validated PCR systems and smaller projects.
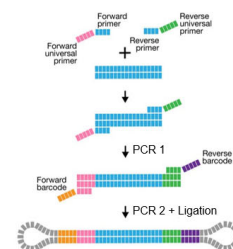   – Besides amplicons, barcoded overhang adapters are also used for microbial multiplexing.

4. **Linear Barcoded Adapters for Probe-Based Captures**
   – Recommended for probe-based captures of genomic regions of interest.
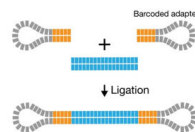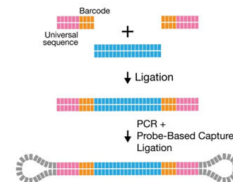

1. Barcoded Target-Specific Primers

2. Barcoded Universal Primer

3. Barcoded Overhang Adapters

4. Barcoded Linear Adapter (Target Capture)

All workflows include barcodes to uniquely identify samples, but the workflows differ in how the barcodes are incorporated into the SMRTbell template.

Description of Linear Barcoded Adapters for Multiplexing Target Capture Experiments

For targeted captures **only**, use Linear Barcoded Adapters:

Linear Barcoded Adapters with Universal Sequences
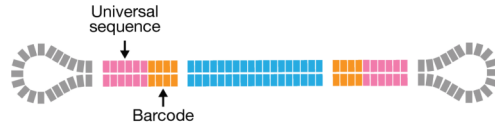
Linear barcoded adapters containing universal sequences are ligated to double-stranded DNA fragments for amplification and capture. This workflow supports up to 24 barcodes, which can be ordered from any oligo synthesis provider.

Demultiplexing of such experiments is supported through SMRT Link command-line options.

Barcode Scoring Modes

The demultiplexing algorithm included in SMRT Link v9.0 only requires information about the barcoding sequences – whether they are the **same** or **different** on either side of the read. Barcode orientation does **not** need to be specified. The demultiplexing algorithm scores **all** barcode sequences in both orientations - forward and reverse complement.

This section describes barcode scoring modes and provides information about the different experiment options depending on barcoding strategy.
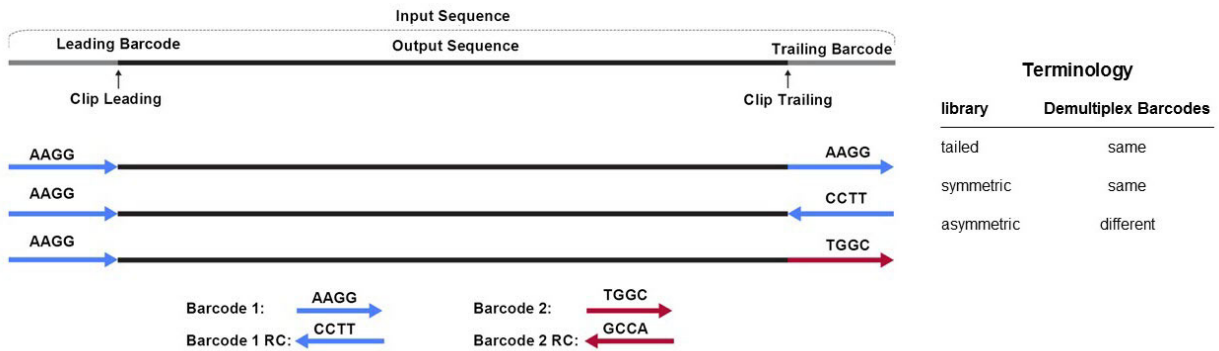
The **Demultiplex Barcodes** application in SMRT Link supports demultiplexing of subreads and CCS data. The following terminology is based on the per (sub-) read view.

Symmetric Mode

For **symmetric** and **tailed** library designs, the **same** barcode is attached to both sides of the insert sequence of interest. The only difference is the orientation of the trailing barcode. For barcode identification, one read with a single barcode region is sufficient. Symmetric barcoding is used for samples constructed using Barcoded Overhang Adapters, Barcoded Universal Primer and target enrichment (linear). This is also the default scoring mode in SMRT Link v9.0 and later.

Asymmetric Mode

Barcode sequences are **different** on the ends of the SMRTbell Template. Asymmetric mode is used with the M13 barcoding procedure. Pacific Biosciences recommends using this mode **only** for small inserts (up to 5 kb) where both ends of the insert are expected to be sequenced. **Both** barcodes must be detected.

Note: For **both** Symmetric and Asymmetric modes, the limit for unique individual barcode sequences is 768, and the limit for the number of different barcode pairs is 15,000.

When running the **Demultiplex Barcodes** applications in SMRT Link, set the **Same Barcodes on Both Ends of the Sequence** option to **Off**.

Mixed Mode

Libraries with combined symmetric and asymmetric barcoding are **not** supported.

Which Barcoding Strategies to Use

For application recommendations, see Overview Sequel Systems Application Options and Sequencing Recommendations.

Note: For multiplexing microbial genomes experiments, click here to see details about the workflows, analysis and barcoding kits that we provide.

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

FASTA File	Sample Preparation		
	Barcoded Adapters	Barcoded Universal Primers	Custom Barcoded Primers
Sequel_16_barcodes_v3	✓[1]		
Sequel_96_barcodes_v1		✓	✓
Sequel_384_barcodes_v1		✓[2]	✓

[1] Using Barcoded Overhang Adapter Kit 8A (PN 101-628-400) and Barcoded Overhang Adapter Kit 8B. (PN 101-628-500)
[2] Using oligos from PacBio recommended lists.

Custom Symmetric Barcode FASTA

Independent of the barcoding strategy and mode, if you are using custom barcodes, you **must** generate a FASTA file of barcode sequences. This file should:

- Include **one** entry per barcode sequence.
- Include **no** duplicate sequences.
- Be orientation-agnostic (forward or reverse-complement, but not reversed.)

```
>bc1000
CTCTACTTACTTACTG
>bc1001
GTCGTATCATCATGTA
>bc1002
AATATACCTATCATTA
```

Please name your barcodes with an alphabetic character prefix to avoid later confusion of barcode name and index. Duplicate names are **not** allowed.

Working with Barcoded Data in SMRT Link v9.0

This section describes how to use SMRT Link to work with barcoded data. **Demultiplex Barcodes** analysis is powered by the Lima SMRT Analysis tool.

The canned data provided with SMRT Link v9.0 includes 7 barcode sets:

- gDNA_Sample_Amplification_Adapter
- IsoSeq_Primers_12_Barcodes_v1
- IsoSeqPrimers_v2 (Includes the content of IsoSeqPrimers as well as support for NEB and Clontech primers.)
- Sequel_16_barcodes_v1
- Sequel_16_barcodes_v3
- Sequel_96_barcodes_v1
- Sequel_384_barcodes_v1

Step 1: Specify the Barcode Setup and Sample Names in a Run Design

1. In SMRT Link, create a new Run Design as described in "Creating a New Run Design" on page 11 of the **SMRT Link User Guide (v9.0)**. **Before** you finish the new Run Design, perform the following steps.

2. Click **Barcoded Sample Options** and then click **Yes** for **Sample is Barcoded**. Additional fields related to barcoding display.
3. Specify a **Barcode Set** using the dropdown list.
   Note: You can specify up to 15,000 samples. Specifying **more** than 10,000 samples may cause a delay of several minutes in analysis submission.
4. Specify if the **same** barcodes are used on both ends of the sequences.
   – Selecting **Yes** specifies symmetric and tailed designs where **all** the reads have the same barcodes on both ends of the insert sequence. Barcode analysis of such experiments retains **only** data with the same barcode identified on both ends.
   – Selecting **No** specifies asymmetric designs where the barcodes are **different** on each end of the insert. Barcode analysis of such experiments retains any barcode pair combination identified in the Data Set.
5. SMRT Link automatically creates a CSV-format **Autofilled Barcode Name** File. The barcode name is populated based on your choice of barcode set, and if the barcodes are the same at both ends of the sequence. The file includes a column of automatically-generated Bio Sample Names 1 through N, corresponding to barcodes 1 through N, for the biological sample names.
   – To specify the biological sample names corresponding to each barcode, click **Download Name File**, enter the biological sample names associated with the barcodes in the second column, and save the file. (Use alphanumeric characters, spaces, hyphens, underscores, colons, or periods **only** - other characters will be removed **automatically,** with a maximum of 40 characters.) If you did **not** use all barcodes in the Autofilled Barcode Name file in the sequencing run, **delete** unused rows.
6. Select the **Barcoded Sample Name File** you edited in **Step 5**.
7. Click **Save**.

Step 2: Perform the Sequencing Run

Load the samples and perform the sequencing run, using the Run Design you created in Step 1. The demultiplexing analysis is performed automatically on the SMRT Link Server once the data is transferred from the Sequel/Sequel II System. This creates an analysis of type `Demultiplex Barcodes (Auto)` in the SMRT Analysis module. You can click to select this analysis and review the reports and data created. If everything looks fine, you can continue to **Step 4** and use the demultiplexed Data Set(s) created by the run as input to further analysis.

Note: By default, `Demultiplex Barcodes (Auto)` creates **one** Data Set per auto-detected barcode within the selected barcode set. It also applies a Data Set filter of a minimum barcode score greater than `26` for optimal results in secondary analyses. If used, the analysis parameter **Filters to add to the Data Set** overrides other barcode filtering, even if the barcode score set with it is lower than `26`.

Step 3: (Optional) Run the Demultiplex Barcodes Application

If you did **not** specify the barcode setup in the Run Design, or if you need to change any of the parameters used in the **Demultiplex Barcodes** analysis automatically launched from Run Design, run the **Demultiplex Barcodes** application. This application separates reads by barcode and creates a new demultiplexed Data Set that you can then use as input to other secondary analysis applications.

1. Click **+ Create New Analysis**.
2. Enter a **name** for the analysis.
3. Select the type of data to use for the analysis:
   - **Sequel Data**: Subreads from Sequel or Sequel II Systems.
   - **CCS Data**: Single Molecule consensus reads generated from the CCS analysis.

   The Data Sets box displays the appropriate Data Sets available for the analysis.
4. In the Data Sets box, select one or more Data Sets to be analyzed together.
5. Click **Next**.
6. Select **Demultiplex Barcodes** from the Applications list.

7. Specify a **Barcode Set** (barcode sequence file.)
   Note: You can specify up to 15,000 samples. Specifying **more** than 10,000 samples may cause a delay of several minutes in analysis submission.
8. Specify the name for the new demultiplexed Data Set that will display in SMRT Link.
9. Specify if the **same** barcodes are used on both ends of the sequences.
   – Selecting **On** specifies symmetric and tailed designs where **all** the reads have the same barcodes on both ends of the insert sequence. Barcode analysis of such experiments retains **only** data with the same barcode identified on both ends.
   – Selecting **Off** specifies asymmetric designs where the barcodes are **different** on each end of the insert. Barcode analysis of such data retains any barcode pair combination identified in the Data Set.
10. Specify the **Minimum Barcode Score**: Reads with barcode scores below the value are **not** included in the analysis. We recommend that you set this value to 26 for **all** applications.
11. SMRT Link automatically creates a CSV-format **Autofilled Barcode Name** File. The barcode name is populated based on your choice of barcode set, and if the barcodes are the same at both ends of the sequence. The file includes a column of automatically-generated Bio Sample Names 1 through N, corresponding to barcodes 1 through N, for the biological sample names.
    – To specify the biological sample names corresponding to each barcode, click **Download Name File**, enter the biological sample names associated with the barcodes in the second column, and save the file. (Use alphanumeric characters, spaces, hyphens, underscores, colons, or periods **only** - other characters will be removed **automatically,** with a maximum of 40 characters.) If you did **not** use all barcodes in the Autofilled Barcode Name file in the sequencing run, **delete** those rows.

– **Note**: Open the CSV file in a text editor and check that the columns are separated by **commas**, not semicolons or tabs.

12. Select the **Barcoded Sample Name File** you edited in **Step 11**.
13. (**Optional**) Specify any advanced parameters.
14. Click **Start**. After the analysis is finished, a new demultiplexed Data Set is available.

**Note**: For information about the reports generated by the Demultiplex Barcodes application, see "Demultiplex Barcodes Reports and Data Files" on page 11.

### Step 4: Run Applications Using the Demultiplexed Data as Input

All secondary analysis applications except **Demultiplex Barcodes** can use demultiplexed Data Sets as input.

Note: For **Iso-Seq** analysis with barcoded samples, use the Iso-Seq application **instead** of the Demultiplex Barcodes application, as the Iso-Seq application already includes the demultiplexing step as part of the pipeline. When performing multiplexed Iso-Seq analysis, ensure that the Run Design **Sample Is Barcoded** option is set to **No** (the default setting). Then, in SMRT Analysis, go straight to the Iso-Seq application and, in the parameters section, select a Primer Set containing multiple primers, such as `IsoSeq_Primers_12_Barcodes_v1`.

1. Select the secondary analysis application to use.
2. Click the number in the **Demultiplexed Subsets** column, then select the demultiplexed Data Set to use as input:
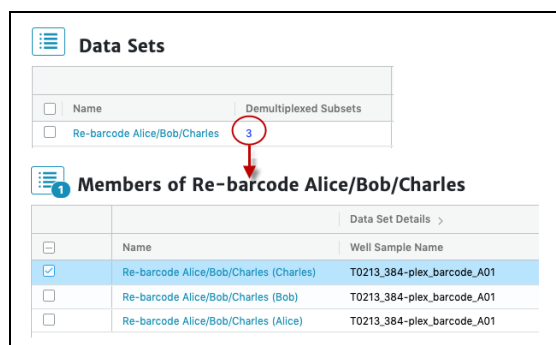
– You can select the **entire** Data Set as input, or one or more specific outputs from selected barcodes, to a maximum of 16 sub-Data Sets, 12 for Iso-Seq Analysis.
3. Additional **Analysis Type** options become available. You can select from the following options:
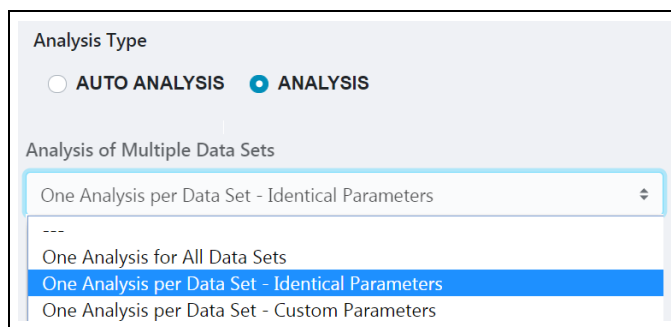
- **One Analysis for All Data Sets**: Runs one analysis using all the selected barcode Data Sets as input, for a maximum of 30 Data Sets.
- **One Analysis per Data Set - Identical Parameters:** Runs **one** separate analysis for **each** of the selected barcode Data Sets, using the **same** parameters, for a maximum of 10,000 Data Sets. Optionally click **Advanced Parameters** and modify parameters.
- **One Analysis per Data Set - Custom Parameters:** Runs **one** separate analysis for **each** of the selected barcode Data Sets, using **different** parameters for each Data Set, for a maximum of 16 Data Sets. Click **Advanced Parameters** and modify parameters. Then click **Start and Create Next**. You can then specify parameters for each of the included barcode Data Sets.
- **Note**: The number of Data Sets listed is based on testing using PacBio's suggested compute configuration, listed in **SMRT Link Software Installation (v9.0)**.

4. Click **Start** to submit the analysis.

Demultiplex Barcodes Input and Output

Given an input set of barcodes and a BAM Data Set, the Demultiplex Barcodes application produces:

- A set of BAM files whose reads are annotated with the barcodes;
- A `subreadset` file that contains the file paths of that collection of barcode-tagged BAM files and their related files.

Barcode Set (Required):

- Specify a barcode sequence file to separate the reads.

New Data Set Name (Required):

- Specify the name for the new demultiplexed Data Set that will display in SMRT Link.

Same Barcodes on Both Ends of Sequence (Default = ON)

- Specify **On** to retain all the reads with the **same** barcodes on both ends of the insert sequence, such as symmetric and tailed designs. (See "Barcode Scoring Modes" on page 3 for information on barcode designs.)
- Specify **Off** to specify asymmetric designs where the barcodes are **different** on each end of the insert.

Minimum Barcode Score (Default = 0)

- A **barcode score** measures the alignment between a barcode attached to a read and an ideal barcode sequence, and is an indicator of how well the chosen barcode pair matches. It ranges between `0` (no match) and `100` (a perfect match). Specifies that reads with barcode scores **below** this minimum value are **not** included in analysis.

Autofilled Barcode Name File

- SMRT Link automatically creates a CSV-format **Autofilled Barcode Name File**. The barcode name is populated based on your choice of barcode set, and if the barcodes are the same at both ends of the sequence. The file includes a column of automatically-generated Bio Sample Names `1` through `N`, corresponding to barcodes `1` through `N`, for the biological sample names.

- To specify the biological sample names corresponding to each barcode, click **Download Name File**, enter the biological sample names associated with the barcodes in the second column, and save the file. (Use alphanumeric characters, spaces, hyphens, underscores, colons, or periods **only** - other characters will be removed **automatically,** with a maximum of 40 characters.) If you did **not** use all barcodes in the Autofilled Barcode Name file in the sequencing run, **delete** those rows.

- **Note**: Open the CSV file in a text editor and check that the columns are separated by **commas**, not semicolons or tabs.

Barcoded Sample File Name

- Select the Barcoded Sample Name File you edited previously.

Demultiplex Barcodes Reports and Data Files

Reports and Data Files

The Demultiplex Barcodes application generates the following reports:

Barcodes > Summary Metrics

- **Unique Barcodes:** The number of unique barcodes in the sequence data.
- **Barcoded Reads:** The number of barcoded reads in the sequence data.
- **Mean Reads:** The mean number of reads per barcode.
- **Max. Reads:** The maximum number of reads per barcode.
- **Min. Reads:** The minimum number of reads per barcode.
- **Mean Read Length:** The mean read length of reads per barcode.
- **Mean Longest Subread Length:** The mean length of the longest subread in each barcoded sample.
- **Unbarcoded Reads:** The number of reads without barcodes in the sequence data.

Barcodes > Barcode Data

- **Bio Sample Name:** The name of the biological sample associated with the barcode.
- **Barcode Index:** The index number associated with the barcode.
- **Barcode Name**: A string containing the pair of barcode indices for which the following metrics apply.
- **Polymerase Reads:** The number of polymerase reads associated with the barcode.

- **Subreads**: The number of subreads associated with the barcode.
- **Bases**: The number of bases associated with the barcode.
- **Mean Read Length:** The mean read length of reads associated with the barcode.
- **Longest Subread Length:** The longest subread length associated with the barcode.
- **Mean Barcode Quality:** The mean barcode quality associated with the barcode.
- **Rank Order (Num. Reads)**: The rank order of this barcode in terms of the number of reads.

Barcodes > Barcoded Read Statistics

- **Number of Reads per Barcode**: Line graph displays the number of sorted reads per barcode.
  - **Good performance**: The Number of Reads per Barcode line (blue) should be mostly linear. Note that this depends on the choice of Y-axis scale. The mean Number of Reads per Barcode line (red) should be near the middle of the graph and should not be skewed by samples with too many or too few barcodes.
  - **Questionable performance**: A sharp discontinuity in the blue line, followed by no yield, with the red line way far from the center. Check the output file **Inferred Barcodes**, note the correct barcodes used, and consider reanalyzing the multiplexed samples with the correct Bio Sample names for the barcodes actually used. If you reanalyze the data, ensure that the **Barcode Name File** includes **only** the correct barcodes used.
- **Barcode Frequency Distribution**: Histogram distribution of read counts per barcode.
  - **Good performance**: A uniform distribution, which is most often a fairly tight symmetric normal distribution, with few barcodes in the tails.
  - **Questionable performance**: A large peak at zero. This can indicate use of incorrect barcodes. Check the output file **Inferred Barcodes**, note the correct barcodes used, and consider reanalyzing the multiplexed samples with the correct Bio Sample names for the barcodes actually used. If you reanalyze the data, ensure that the **Barcode Name File** includes **only** the correct barcodes used.
- **Mean Read Length Distribution**: Histogram distribution of the mean polymerase read length for all samples.
  - **Good performance**: The distribution should be normal with a relatively tight range.
  - **Questionable performance**: A spread out distribution, with a mode towards the low end.

Barcodes > Barcode Quality Scores

- **Barcode Quality Score Distribution:** Histogram distribution of barcode Quality scores. The scores range from 0-100, with 100 being a perfect match. Any significant modes or accumulation of scores <40 suggests issues with some of the barcode analyses. The red line is set at 26 – the minimum default barcode score.
  - **Good performance**: Distributions with a mode >65 and the low-end tail tapering off below 40.

Barcode Quality Score Distribution
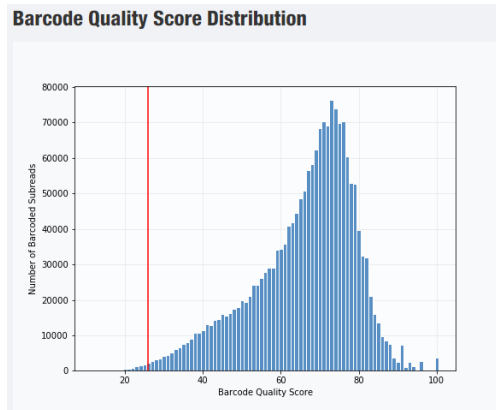
- **Questionable performance**: A bimodal distribution with a large second peak usually indicates that some barcodes that were sequenced were **not** included in the barcode scoring set.

Barcodes > Barcoded Read Binned Histograms

- **Read Length Distribution By Barcode**: Histogram distribution of the Polymerase read length by barcode. Each column of rectangles is similar to a read length histogram rotated vertically, seen from the top. Each sample should have similar Polymerase read length distribution. Non-smooth changes in the pattern looking from left to right might indicate suboptimal performance.
- **Barcode Quality Distribution By Barcode**: Histogram distribution of the per-barcode version of the **Read Length Distribution by Barcode** histogram. The histogram should contain a single cluster of hot spots in each column. All barcodes should also have similar profiles; significant differences in the pattern moving from left to right might indicate suboptimal performance.
    - **Good performance**: All columns show a single cluster of hot spots.
    - **Questionable performance**: A bimodal distribution would indicate missing barcodes in the scoring set.

Data > File Downloads

The following files are available on the Analysis Results page. Additional files are available on the SMRT Link server, in the analysis output directory.

- **Analysis Log**: Log information for the analysis execution.
- **SMRT Link Log**: Server-level analysis log information. (This file is displayed when you choose **Data > SMRT Link Log**.)
- **Barcode Files**: Barcoded subread Data Sets; one file per barcode.
- **Barcode Summary CSV**: Data displayed in the reports, in CSV format. This includes Bio Sample Name.
- **Barcode Summary:** Text file listing how many ZMWs were filtered, how many ZMWs are the same or different, and how many reads were filtered.
- **Inferred Barcodes**: Inferred barcodes used in the analysis. The barcoding algorithm looks at the first 35,000 ZMWs, then selects barcodes with ≥10 counts and mean scores ≥45.
- **Unbarcoded Reads**: BAM file containing reads not associated with a barcode.

Note: You can get the demultiplexed BAM files using the Data Management module's "Export Data Sets" feature. In the demultiplexed

BAM output file a tag `bc` is added for each read, indicating the assigned barcode. The `bc` tag is the zero-based index of the barcodes in the FASTA file. For example, when using the barcodes `RSII_96_barcodes`, a subread with barcode `lbc1` identified on both sides will have the tag `bc:B:S,0,0` in the BAM output file.

A second `bq` tag corresponds to the barcode quality (0-100).

For details on running the Demultiplex Barcodes application using the command-line, see the document **SMRT Tools Reference Guide (v9.0)**, available here.

Working with Demultiplexing Results

Finding Demultiplexed subreads.bam Files

Subreads.bam files **cannot** be downloaded from the SMRT Analysis Report page. You can find these files using the command line by navigating to the SMRT Link server. To do so:

1. Copy the job path (**Analysis Overview > Status > Path**) on the SMRT Link server shown on the Report page.

2. Open the command line, login into the SMRT Link Server and change directory to the path you copied:

```
$ cd /pbi/analysis/smrtlink/release/smrtsuite/userdata/jobs_root/0000/0000029/
0000029399
```

3. Display the folders and files in the directory of the Barcode Demultiplexing Run:

```
$ ls -l
lrwxrwxrwx 1 fas Domain Users  140 Oct 10 09:36 cromwell-job -> /pbi/analysis/smrtlink/
release/smrtsuite/userdata/jobs_root.local/cromwell-executions/pb_demux_subreads/
c9cf0a1c-5fef-46d9-97e4-3ad44d80b597
drwxr-xr-x 2 fas Domain Users  4096 Oct 10 09:36 entry-points
drwxr-xr-x 3 fas Domain Users  4096 Oct 10 10:04 logs
drwxr-xr-x 2 fas Domain Users 20480 Oct 10 10:04 outputs
-rw-r--r-- 1 fas Domain Users     0 Oct 10 09:36 pbscala-job.stderr
-rw-r--r-- 1 fas Domain Users 22778 Oct 10 10:05 pbscala-job.stdout
drwxr-xr-x 2 fas Domain Users  4096 Oct 10 10:04 workflow
```

Notice that the `cromwell` job is a hyperlink to another directory, designed to save space on your SMRT Link server.

4. Navigate to the subdirectory where the .bam files are stored:

```
$cd cromwell-job/call-demultiplex_barcodes/demultiplex_barcodes/083b56b3-dee2-44ae-aa97-e6656bb4763f/call-lima/execution

$ls -l

-rw-r--r-- 1 fas Domain Users 215561027 Oct 10 09:58 lima.bc1001--bc1001.bam
-rw-r--r-- 1 fas Domain Users    388923 Oct 10 09:58 lima.bc1001--bc1001.bam.pbi
-rw-r--r-- 1 fas Domain Users     12857 Oct 10 09:58 lima.bc1001--bc1001.subreadset.xml
-rw-r--r-- 1 fas Domain Users 227808959 Oct 10 09:58 lima.bc1002--bc1002.bam
-rw-r--r-- 1 fas Domain Users    408837 Oct 10 09:58 lima.bc1002--bc1002.bam.pbi
-rw-r--r-- 1 fas Domain Users     12857 Oct 10 09:58 lima.bc1002--bc1002.subreadset.xml
-rw-r--r-- 1 fas Domain Users 217156999 Oct 10 09:58 lima.bc1003--bc1003.bam
-rw-r--r-- 1 fas Domain Users    390196 Oct 10 09:58 lima.bc1003--bc1003.bam.pbi
-rw-r--r-- 1 fas Domain Users     12857 Oct 10 09:58 lima.bc1003--bc1003.subreadset.xml
-rw-r--r-- 1 fas Domain Users 238209137 Oct 10 09:58 lima.bc1004--bc1004.bam
-rw-r--r-- 1 fas Domain Users    428762 Oct 10 09:58 lima.bc1004--bc1004.bam.pbi


-rw-r--r-- 1 fas Domain Users     43595 Oct 10 09:58 lima.json
-rw-r--r-- 1 fas Domain Users      2757 Oct 10 09:58 lima.lima.counts
-rw-r--r-- 1 fas Domain Users      2977 Oct 10 09:41 lima.lima.guess
-rw-r--r-- 1 fas Domain Users 433394971 Oct 10 09:58 lima.lima.report
-rw-r--r-- 1 fas Domain Users       791 Oct 10 09:58 lima.lima.summary
-rw-r--r-- 1 fas Domain Users 992123522 Oct 10 09:58 lima.removed.bam
-rw-r--r-- 1 fas Domain Users   2413108 Oct 10 09:58 lima.removed.bam.pbi
-rw-r--r-- 1 fas Domain Users       569 Oct 10 09:58 lima.removed.json
-rw-r--r-- 1 fas Domain Users     12542 Oct 10 09:58 lima.removed.subreadset.xml
-rw-r--r-- 1 fas Domain Users         2 Oct 10 09:58 rc
-rw-r--r-- 1 fas Domain Users      4459 Oct 10 09:37 script
-rw-r--r-- 1 fas Domain Users      1162 Oct 10 09:37 script.submit
-rw-r--r-- 1 fas Domain Users         0 Oct 10 09:37 stderr
-rw-r--r-- 1 fas Domain Users         0 Oct 10 09:37 stderr.submit
-rw-r--r-- 1 fas Domain Users         0 Oct 10 09:37 stdout
-rw-r--r-- 1 fas Domain Users        63 Oct 10 09:37 stdout.submit
```

For **each** demultiplexed barcode, the Barcode Demultiplexing workflow generates three files: `subread.bam`, `subread.bam.pbi` and `subreadset.xml`.

Other files are generated and display in the advanced report:

`lima.lima.counts, lima.lima.guess, lima.lima.report, lima.lima.summary,lima.removed.bam`

The json files (`lima.json` and `lima.remove.json`) are useful for running a user-defined workflow.

Note: For information on working with barcoded data in `pbcromwell`, see page 62 of **SMRT Tools Reference Guide (v9.0)**.

R plotting scripts are available on github:
https://github.com/pacificbiosciences/barcoding/#quality-control.

For information on working with barcoded data in lima, see:
https://github.com/PacificBiosciences/barcoding.

Additional Information:

- Demultiplex Barcodes Algorithm:
  https://github.com/PacificBiosciences/barcoding
- PacBio BAM Format Specifications:
  http://pacbiofileformats.readthedocs.io/en/9.0/BAM.html