SMRT® Link Getting Started Guide (v4.0.0)

Introduction

This document describes how to use Pacific Biosciences' SMRT® Link software to:

1. Design sequencing runs and create and/or import sample sheets using the **Run Design** module.
2. Monitor run progress, status and quality metrics using the **Run QC** module.
3. Create Projects and Data Sets; generate QC reports for Data Sets; and view or import sequence data using the **Data Management** module.
4. Perform **secondary analysis** on the basecalled data (such as *de novo* assembly, variant detection, alignment, or SNP identification) after the run has completed, using the **SMRT Analysis** module.
5. **Visualize** data using **SMRT View**, a genome browser that displays sequencing data generated by the PacBio RS II and Sequel™ Systems.

This document also describes:

- The data generated by the Sequel System for each cell that are transferred to network storage.
- Configuration and user management.

PacBio RS II System Users

SMRT Link's **Data Management** and **SMRT Analysis** modules are compatible with PacBio RS II data; the rest of the SMRT Link modules work **only** with Sequel Systems.

- PacBio RS II users should use **RS Remote**, and **RS Dashboard** software instead of the SMRT Link **Run Design**, and **Run QC** modules.
- To display **only** the modules for the PacBio RS II: Choose **Configure** from the SMRT Link menu, click **Instrument**, then check the **PacBio RS II Only** box. SMRT Link displays only the **Data Management** and **SMRT Analysis** modules.

Contact Information

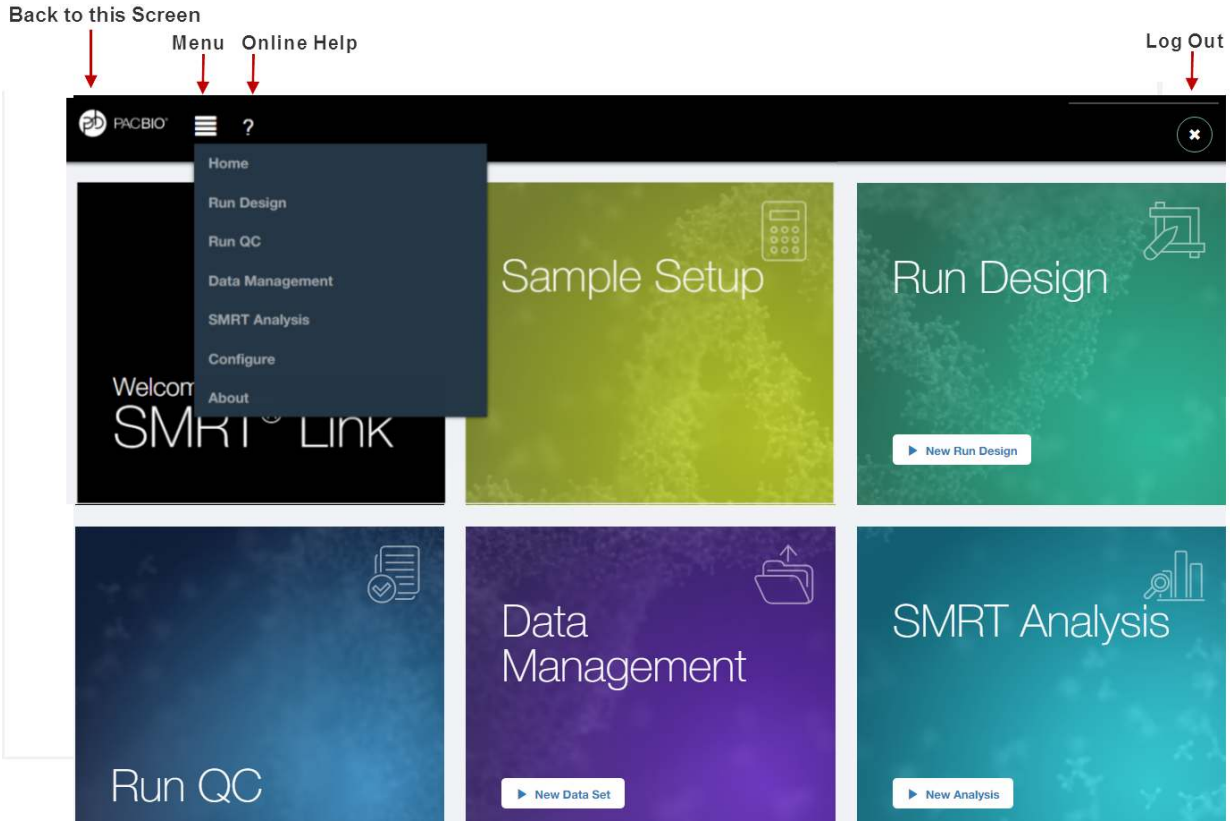
For additional technical support, contact Pacific Biosciences at TechSupport@pacb.com or 877 920-PACB.

Using SMRT® Link

You access SMRT Link using the Chrome web browser.

- SMRT Link is **not** available on the instrument – it must be accessed from a remote workstation.
- Depending on how SMRT Link was installed at your site, logging in with a user name and password may be required.
- SMRT Link needs a Secure Sockets Layer (SSL) Certificate to ensure a secure connection between the SMRT Link server and your browser using the HTTPS protocol.
- If an SSL Certificate is **not** installed with SMRT Link, the application will use the PacBio self-signed SSL Certificate and will use the HTTP protocol. In this case, **each** user will need to accept the browser security warnings described in “If Using the PacBio Self-Signed SSL Certificate” on page 23.

After accessing SMRT Link, the **Home** page displays. (Shown is the interface for the Sequel System.)



- Click the **menu** to navigate to any of the modules, configure for the PacBio RS II, view version information, or perform administrative functions (Admins **only**).
- Click a module **name** to access that module. **Run Design**, **Data Management** and **SMRT Analysis** include links to create new Run Designs, Data Sets, and Analyses.

- Click the **PacBio logo** at the top left to navigate back to the SMRT Link Home page from within the application.
- Click **?** to view the SMRT Link Online help.
- Click **X** to log out of SMRT Link.

- Within a module: Click the **module name** or the **module design** to navigate back to the module's home screen.

Working with tables

- To **sort** table columns: Click a **column title**.
- To **search** within a table: Enter a unique search term into the **Search** field.
- To **show or hide** table fields: Click the control, then check or uncheck column name(s) in the dialog box that displays.

Creating Run Designs

Use SMRT Link's **Run Design** module to create, edit, or import Run Designs. A **Run Design** specifies:

- The wells and SMRT® Cells to include in the sequencing run.
- The collection and analysis applications to use for the selected wells and cells.

The Run Design then becomes available from **Sequel™ Instrument Control Software (ICS)**, the instrument touchscreen software used to select a Run Design, load the instrument, and then start the run.

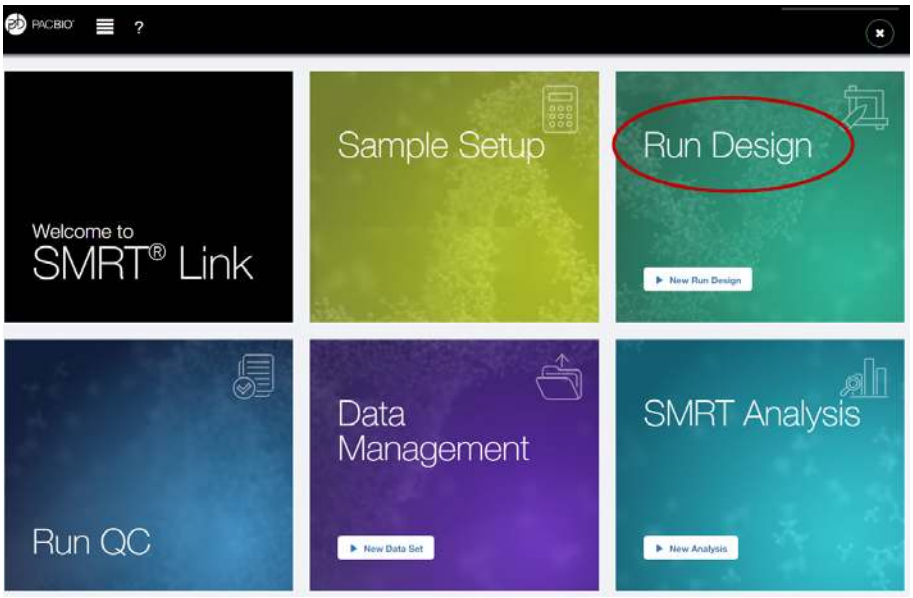
Run Designs created in SMRT Link are accessible from **all** Sequel Systems linked to SMRT Link.

- If you are using SMRT Link with a PacBio RS II, use **RS Remote**.

SMRT Link includes two different ways to create a Run Design:

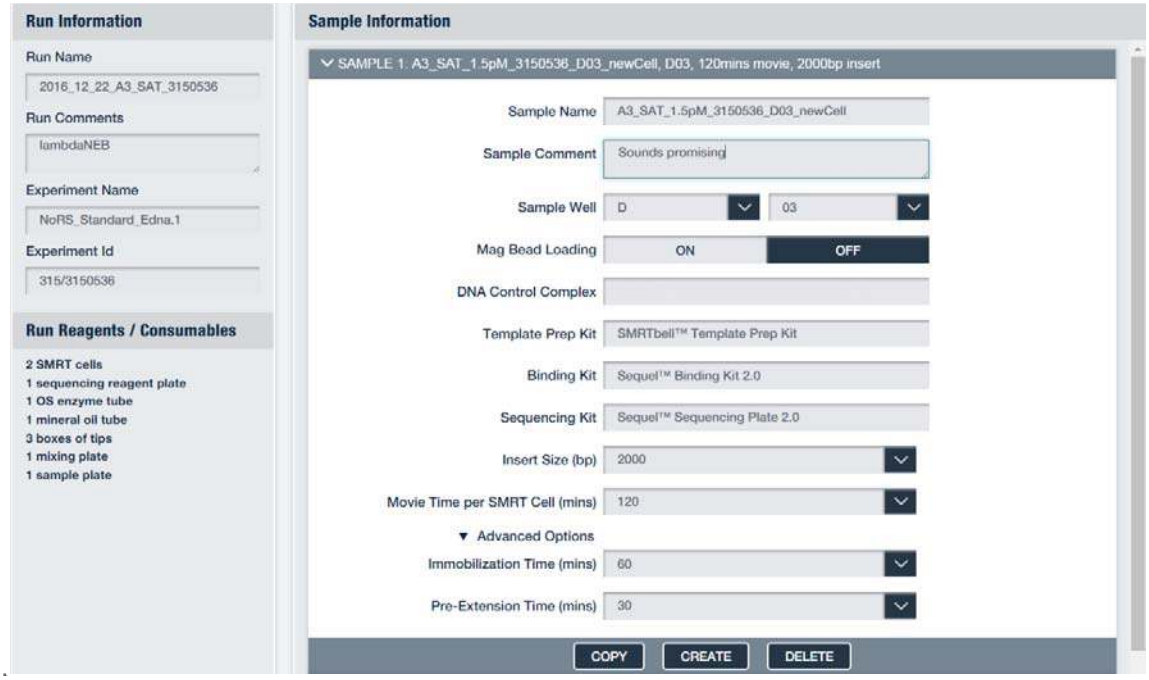
1. Use SMRT Link's **Run Design** module to create a new run.
2. Create a CSV file, then import it using SMRT Link's **Run Design** module.

Creating a New Run Design in SMRT Link



1. Access SMRT Link using the Chrome web browser.
2. Select **Run Design**.
3. Runs Designs can be sorted and searched for:
   - To sort Run Designs, click a **column title**.
   - To search for a Run Design, enter a unique search term into the **Search** field.

4. Click **New Run Design**.



5. Enter a **Run Name**. (The software creates a new Run Name based on the current date and time; you can edit the name as needed.)
6. (**Optional**): Enter **Run Comments**, **Experiment Name**, and **Experiment ID** as needed. (**Note**: Experiment ID **must** be alphanumeric.)
7. Next, specify information for the first sample that will be part of the run. Enter a **Sample Name**.
8. (**Optional**) Enter **Sample Comments**.
9. Specify the **well position** used for this sample.
10. Specify whether or not to use **MagBead** Loading. (This immobilizes SMRTbell™ templates into the ZMWs on the SMRT Cell using Mag-Beads. Pacific Biosciences **highly recommends** using MagBead for SMRTbell templates with inserts >1 kb. SMRTbell templates >1 kb load better and have a higher raw accuracy compared to diffusion loading.)
    When **OFF** is selected, a diffusion run will be performed.
11. (**Optional**) Specify a DNA Control Complex. (A DNA Control Complex is **not** yet available at this time.)
12. Enter or scan a **Template Prep Kit** barcode number. The Template Prep Kit name displays. If the barcode is invalid, "Invalid barcode" displays.
13. Enter or scan a **Binding Kit** barcode number. The Binding Kit name displays. If the barcode is invalid, "Invalid barcode" displays.
14. Enter or scan a **Sequencing Kit** barcode. The barcode insert is in the Sequencing Kit box. Enter the barcode by scanning the insert. The Sequencing Kit name displays.

- **Note**: If the Sequencing or Binding kit is **incompatible**, an error message displays indicating the obsolete chemistry, and the run is **prevented** from proceeding.

15. Specify an **Insert Size**, ranging from 500 to 20,000 base pairs. (The Insert Size is the length of the double-stranded nucleic acid fragment in a SMRTbell template, excluding the hairpin adapters. This matches the average insert size for the sample; the size range boundaries are described in the library preparation protocol and in the table of loading recommendations.)
16. Specify the **Movie time** (**collection time)** per SMRT Cell: 30, 120, 240 or 360 minutes.
17. (**Optional**) Click **Advanced Options**, then specify the length of time (60, 120 or 240 minutes) for **immobilization** of SMRTbell templates.
    - For **MagBead Loading**, this is the length of time the SMRT Cell is at the MagBead station and the magnet moves the MagBead-bound SMRTbell template across the SMRT Cell to deposit into the ZMWs.
    - For **diffusion**, this is the length of time the SMRT Cell is at the Cell Prep Station to allow diffusion of SMRTbell templates to occur into the ZMWs.
    - Default immobilization times are 60 minutes for Diffusion Loading and 120 minutes for MagBead Loading.
18. (**Optional**) Click **Advanced Options**, then specify the length of pre-extension time. This initiates the sequencing reaction prior to data acquisition. After the specified time, the sequencing reagents are removed and replenished with fresh reagents, and data acquisition starts. This feature is targeted at short inserts (such as amplicons <5 kb) and will provide a significant increase in read length. **Note**: This is **not** compatible with Sequencing Kit v1.2 and v1.2.1. If these are used, the run will abort.
19. Sample options:
    - Click **Copy**. This starts a new sample, using the values entered in the first sample.
    - Click **Create**. This starts a new, empty sample.
    - Click **Delete**. This deletes the current sample.
20. After filling in all the samples, click **Save** - this saves the entire Run Design. The new Run Design displays on the main Run Design page.
21. Click **View Summary** to view a table summarizing the entire Run Design. The Run Design file is now imported and available for selection in Sequel ICS on the instrument.

Creating a Run Design by Importing a CSV file

On a remote workstation, open the sample CSV file included with the installation. (Contact your FAS for a copy of the file if it is not available.)

1. Update the file as necessary for the Run Design. (See the definitions of the Run Design attributes in the table below.)
2. Save the edited CSV file.

Field	Required	Description
Run Name	Yes	Run name **must** be entered for the first cell and will be applied to the remaining cells in the run. Use only alphanumeric characters.
Run Description	No	Run description must be entered for first cell and will be applied to the remaining cells in the run. No character limitations; use only alphanumeric characters.
Well No.	Yes	Format as follows: A01, B12. Not acceptable: A1,B3 and so on. Use only alphanumeric characters.
Sample Name	Yes	Enter Sample name. No character limitations; use only alphanumeric characters.
Sample Description	No	Enter Sample description. No character limitations; use only alphanumeric characters.
Insert Size	Yes	Enter values rounded to the nearest kb. (Example: 1000, 2000, 10000, and so on.)
Size Selection	Yes	Enter "Yes" or "No".
Automation Name	Yes	Enter "Magbead" or "Diffusion".
Collection Time	Yes	Enter the desired length of time for the cell run, in minutes, between 30 and 240.
DNA Template Prep Kit Box Barcode	Yes	Enter or scan the template prep barcode. **Working example:** DM1117100259100111716
DNA Control Complex Box Barcode	No	Enter or scan the control DNA barcode. (Not available at this time.)
Binding Kit Box Barcode	Yes	Enter or scan the Binding Kit barcode. **Working example:** DM1117100619300111716
Sequencing Kit Box Barcode	Yes	Enter or scan the Sequencing Kit barcode. **Working example:** DM1117100620000111716
Automation Parameters	No	To enable Pre-Extension time, enter the number of minutes. Example 90 minutes: "ExtensionTime=double:90|ExtendFirst=boolean:True" (**Note**: Leave blank when **not** using pre-extension time.)

Importing a Run Design

After creating the CSV file, import the file into Sequel ICS using SMRT Link.

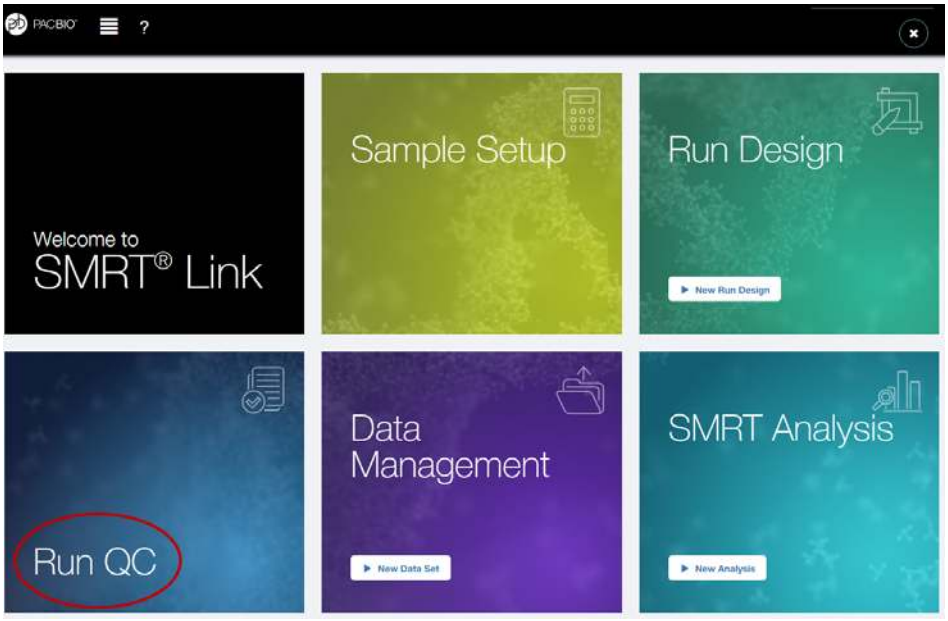
1. Access SMRT Link using the Chrome web browser.
2. Select **Run Design**.
3. Click **Import Run Design**.
4. Select the saved CSV file designed for the run and click **Open**. The file is now imported and available for selection in Sequel ICS on the instrument.

Reviewing Run QC

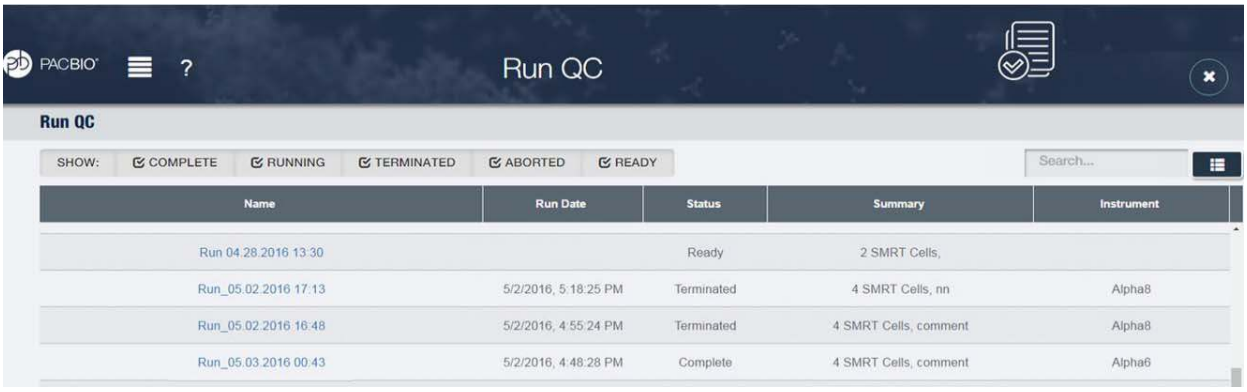
After a run has completed, use SMRT Link's **Run QC** module to monitor performance trends and perform run QC remotely.

Metrics can be reviewed in the Run QC module. **All** Sequel Systems connected to SMRT Link can be reviewed using Run QC.

- If you are using SMRT Link with a PacBio RS II, use **RS Dashboard**.

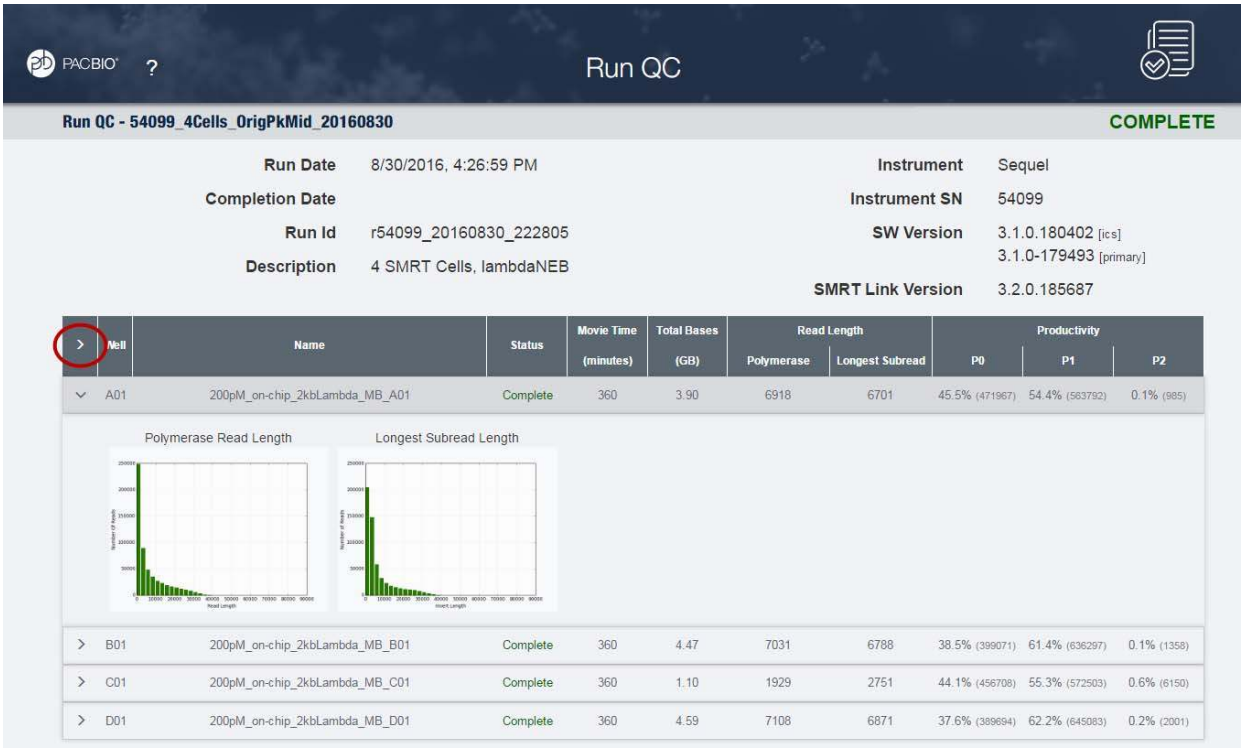


1. Access SMRT Link using the Chrome web browser.
2. Select **Run QC**.



3. Runs can be sorted, searched for, and filtered:
   - To sort runs, click a **column title**.
   - To search for a run, enter a unique search term into the **Search** field.
   - To specify the status of the runs to display, click one or more of the following buttons: **Complete**, **Running**, **Terminated**, **Aborted**, **Paused**, and/or **Ready**.

4. Click the **Run name** of interest. Following are definitions of some of the metrics displayed:
   - **Total Bases**: Sum all of bases from high quality reads.
   - **Polymerase Read Length**: Mean polymerase read length.
   - **Longest Subread Length**: Mean of the maximum subread length per ZMW.
   - **P0**: Empty ZMW; no high quality read detected.
   - **P1**: ZMW with a high quality read detected.
   - **P2**: Other, signal detected but no high quality read.



5. Click the **>** arrow to expand rows to view histograms for each SMRT Cell where data was successfully transferred.

Managing Data Sets and Projects

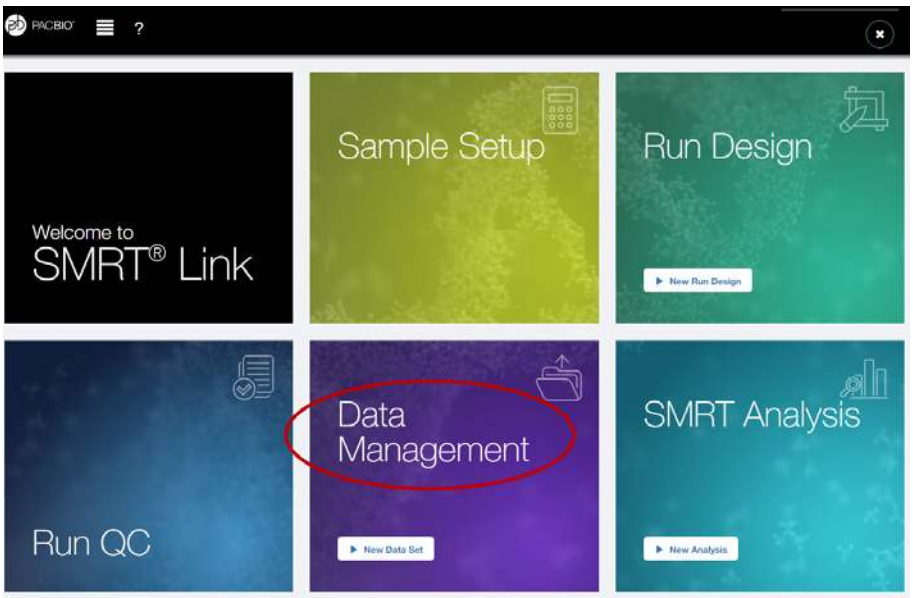
What is a Data Set?

Data Sets are logical collections of sequencing data (raw or analyzed) that are analyzed together, and for which reports are created.

Data Sets:

- Help to **organize** and **manage** raw and analyzed data. This is especially valuable when dealing with large amounts of data collected from different sequencing runs from one or more instruments.
- Are the way that sequence data is represented and manipulated in SMRT Link. Sequence data from the instrument is organized in Data Sets. Data from **each** cell or collection is a Data Set.

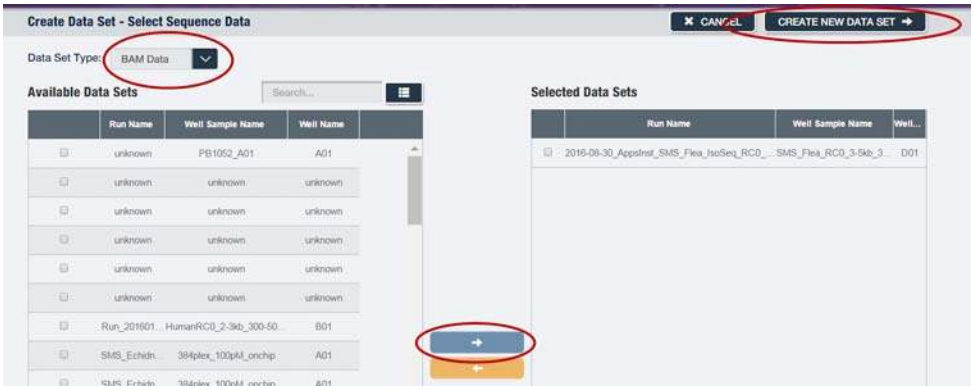
Creating a Data Set



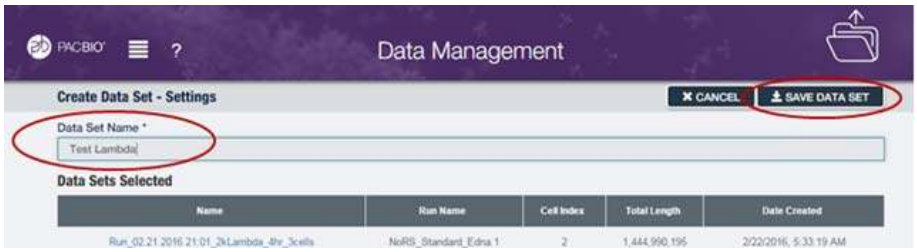
1. Access SMRT Link using the Chrome web browser.
2. Select **Data Management**.
3. Data Sets and Projects can be sorted and searched for:
   - To sort Data Sets or Projects, click a **column title**.
   - To search for a Data Set or Project, enter a unique search term into the **Search** field.

4. Click **+ Create Data Set**.
5. Select the type of data to include in the new Data Set: **BAM-format** data generated on the Sequel System, or **PacBio® RS II** data.

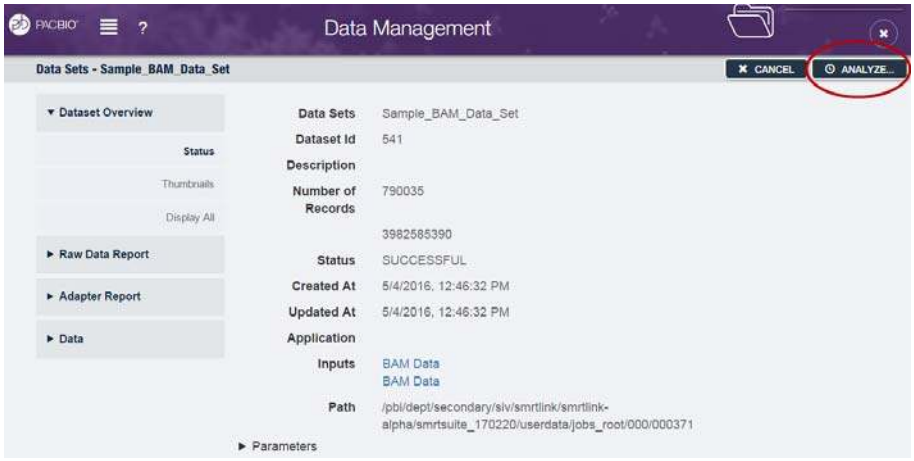
6. In the **Available Data Sets** box, select one or more sets of sequence data.
7. Click the **Right Arrow** button to add the selected sequence data to the new Data Set, then click **Create New Data Set**.



8. Enter a name for the new Data Set, then click **Save Data Set**. The new Data Set becomes available for starting analyses, viewing, or generating reports.



9. After the Data Set is created, click its name in the main Data Management screen to see reports, metrics, and charts describing the data included in the Data Set.



Starting an Analysis from a Data Set

From the Data Set page, an analysis can be started using the Data Set.

1. Click **Analyze**, then follow the instructions starting at Step 5 of "Creating and Starting Analyses" on page 15.

Note: To analyze PacBio RS II data, first convert it using the **Convert RS to BAM** application. After converting the data to the BAM file format, analyzing the data using **all** the other applications is enabled.

What is a Project?

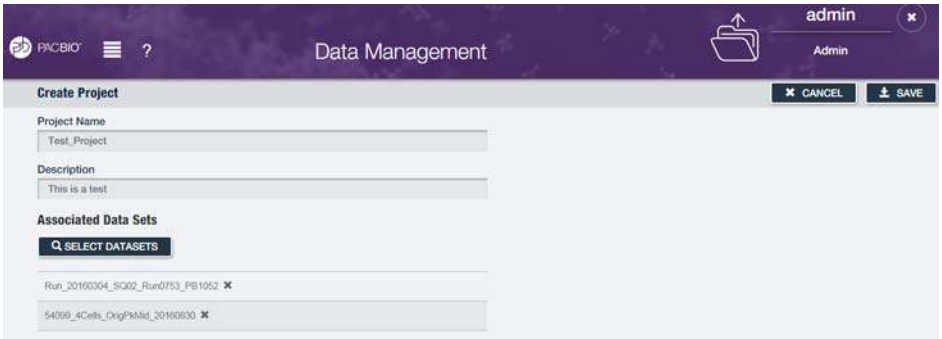
- Projects are collections of Data Sets (raw or analyzed).
- By default, all Data Sets and data belong to the **General** project and are accessible to all users of SMRT Link.

What's in a Project?

- One or more Data Sets and Quality Control information.
- One or more Analysis results, including information for all analysis parameters and reference sequence (if used).

Creating a Project

1. Access SMRT Link using the Chrome web browser.
2. Select **Data Management**.
3. Click **+ Create Project**.
4. Enter a name for the new project.
5. (Optional) Enter a description for the project.
6. Click **Select Data Sets** and select one or more sets of sequence data to associate with the project.
7. Click **Save**. The new project becomes available.



Editing a Project

1. On the Home Page, select **Data Management**.
2. Click **View > Projects**.
3. Click the name of the project to edit.
   - (Optional) Edit the project name or description.
   - (Optional) Delete a Data Set associated with the project by clicking **X**.

- (Optional) Add one or more sets of sequence data to the project by clicking **Select Data Sets** and selecting one or more Data Sets to add.

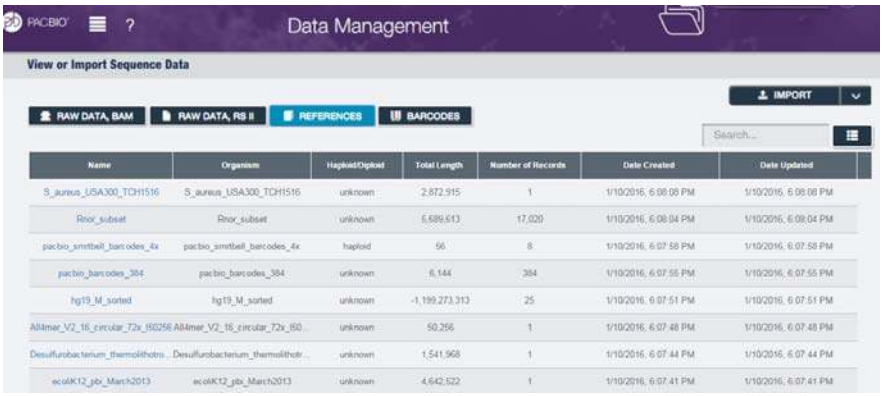
Deleting a Project

1. On the Home Page, select **Data Management**.
2. Click **View > Projects**.
3. Click the name of the project to delete.
4. Click **Delete**.

Viewing Reference and Barcode Data

On the **View or Import Sequence Data** page, information on available reference sequence files and barcode files can be accessed. Note that this data can only be **viewed**, not modified.

1. On the Home Page, select **Data Management**.
2. Click **View or Import Sequence Data**.
3. Click **References** or **Barcodes** to specify what type of data to view. A list of available reference sequence files or barcode files displays.



4. In the **Name** column, click the name of the Reference or Barcode of interest. Details for that reference sequence file or barcode file display.

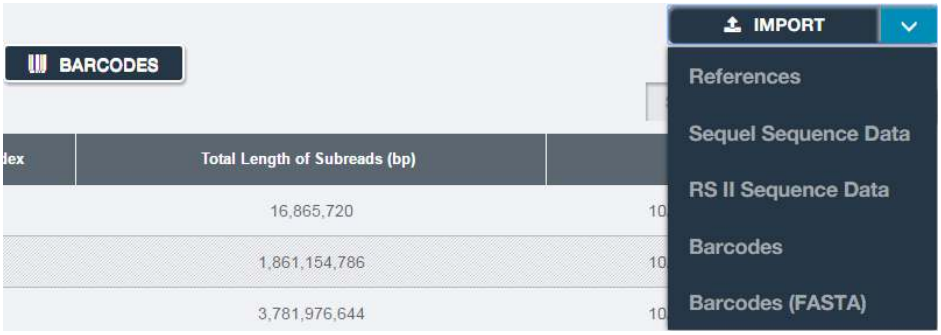
Importing Data

Note: If your Sequel instrument is linked to the SMRT Link software during the instrument installation, your Sequel System data will be **automatically** imported in SMRT Link.

Three types of sequence data, as well as Barcode files, can be imported for use in SMRT Link:

- **Reference sequence files** - FASTA files containing a reference sequence for use in starting analyses.
- **Sequel sequence data** - A file (`.subreads.xml`) file containing information about Sequel sequence data, such as paths to the BAM files.
- **RS II sequence data** - A file (`.metadata.xml`)containing information about PacBio RS II sequence data from a single cell.
- **Barcodes/Barcodes (FASTA)** - `.xml` or FASTA-format files containing barcodes.)

1. On the Home Page, select **Data Management**.
2. Click **View or Import Sequence Data**.
3. Click **Import**, then select either **References**, **Sequel Sequence Data**, **RS II Sequence Data, Barcodes,** or **Barcodes (FASTA)**.

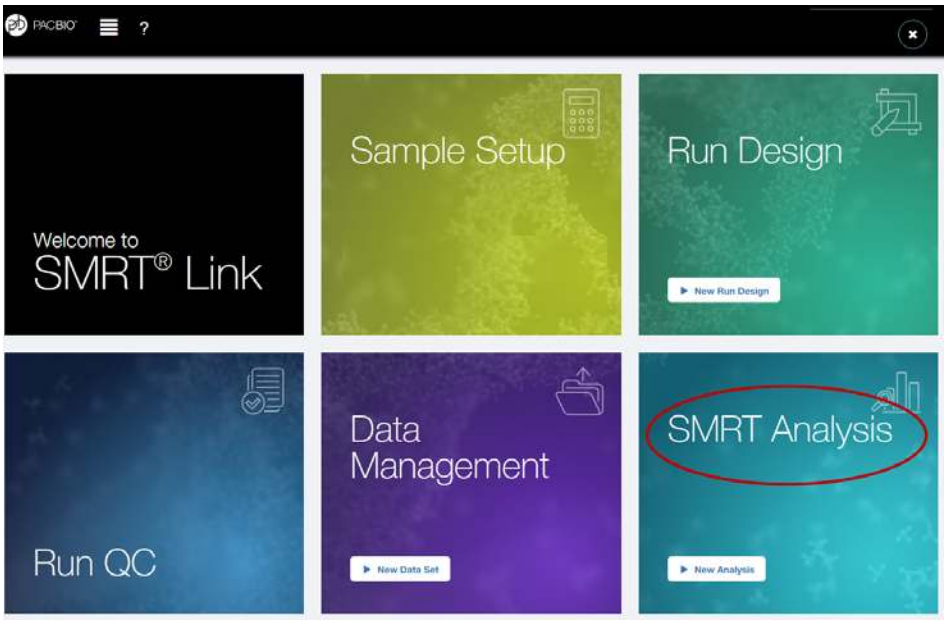


4. Navigate to the appropriate file and click **Import**. The sequence data or barcodes are imported and becomes available in SMRT Link.

For additional help on working with Data Sets and Projects, or viewing and importing sequence data, see the **SMRT Link Online Help** by clicking the **?** icon in the upper-left corner of the screen.

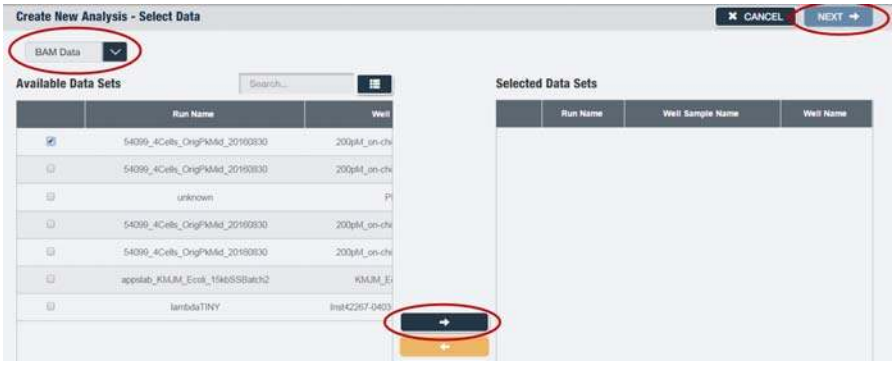
Creating and Starting Analyses

After a run has completed, use SMRT Link's **SMRT Analysis** module to perform **secondary analysis** of the data.



1. Access SMRT Link using the Chrome web browser.
2. Select **SMRT Analysis**.
3. Analyses can be sorted, searched for, and filtered:
   - To sort analyses, click a **column title**.
   - To search for a analysis, enter a unique search term into the **Search** field.
   - To specify the status of the analyses to display, click one or more of the following buttons: **Created**, **Running**, **Submitted**, **Terminated**, **Successful**, and/or **Failed**.

4. Click **+ Create New Analysis**.
5. Specify whether to use **BAM-format** data generated on the Sequel System, or **PacBio RS II** data.
6. In the **Available Data Sets** box, select one or more sets of data to be analyzed together.
7. Click the **Right Arrow** button to add the selected sequence data for analysis, then click **Next**.

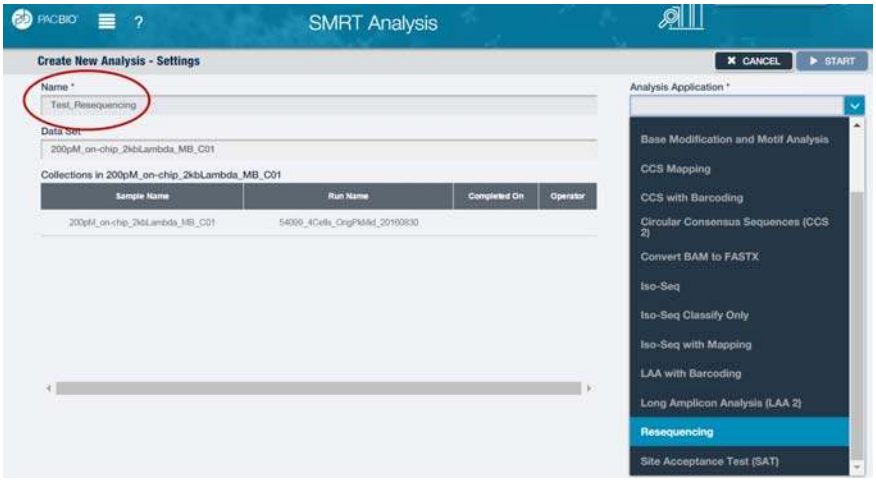


8. Enter a **name** for the analysis.
9. Select a secondary analysis **application** to use for the analysis.

Note: If PacBio RS II data was selected for the analysis, the **only** application available is **Convert RS to BAM**. After converting the data to the BAM file format, analyzing the data using **all** the other analysis applications is enabled.

- Each of the secondary analysis applications has **required parameters** that are displayed. Please review the default values shown.
- Secondary analysis applications also have **advanced analysis parameters**. These are set to default values, and need only be changed when analyzing data generated in non-standard experimental conditions.

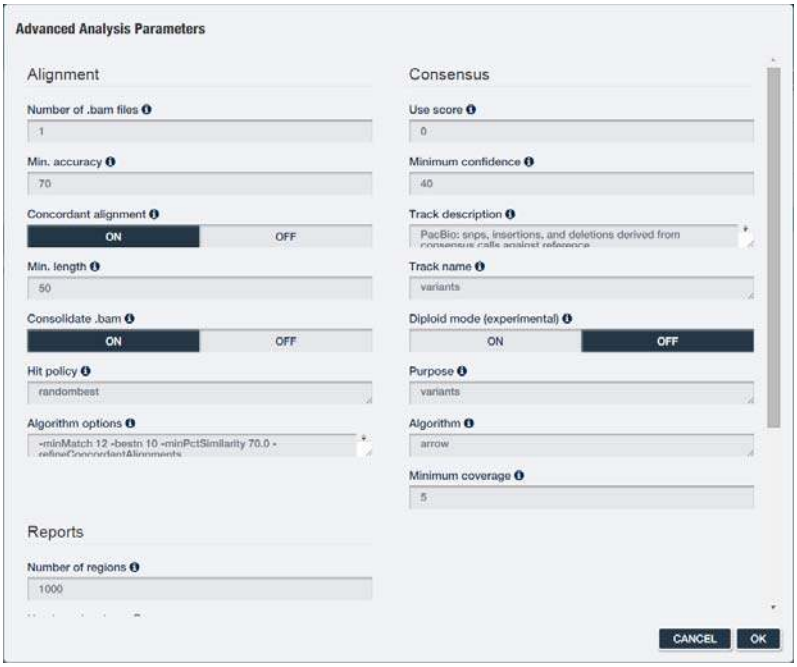
The **Resequencing** application will be used as an example. This application maps sequencing reads against a reference sequence. It identifies the consensus sequence and performs variant detection using the Arrow algorithm.



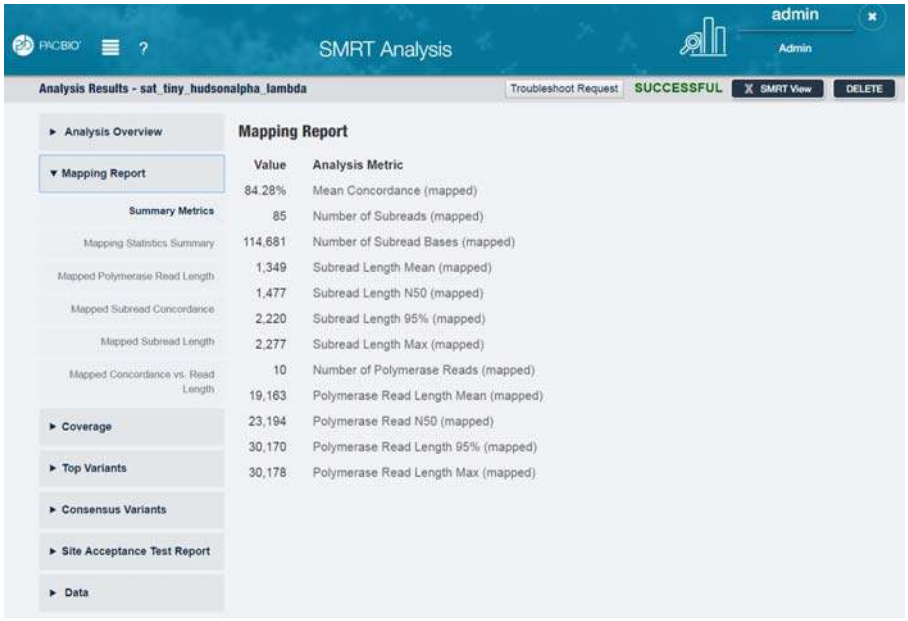
10. Click the **References** field and select a reference sequence from the dialog. (The reference sequences available in SMRT Link and dis-

played in the dialog were imported into SMRT Analysis. See “Importing Data” on page 14 for details.)

11. **(Optional)** Click **Advanced Analysis Parameters** and specify the values of the parameters you would like to change. Click **OK** when finished.



12. Click **Start**.
13. Click **SMRT Analysis** to navigate to the main SMRT Analysis screen. There, the status of the analysis displays. When the analysis has **completed**, click on its name - reports are available for the completed analysis.
14. (Optional) To **delete** the completed analysis: Click **Delete**, then click **Yes** in the confirmation dialog. The analysis is deleted.

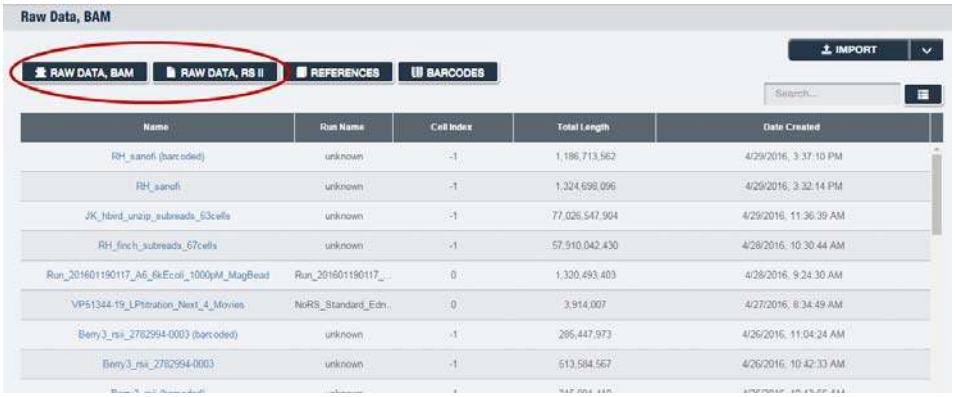


For additional help on starting analyses, see the **SMRT Link Online Help** by clicking the **?** icon in the upper-left corner of the screen.

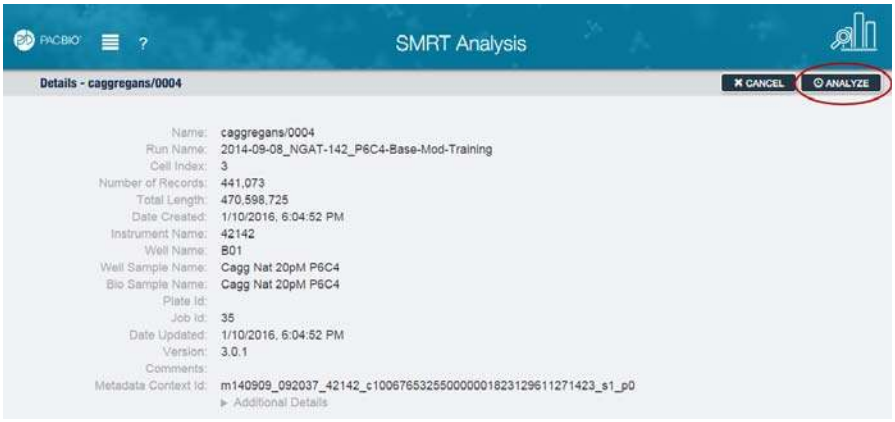
Starting Analyses After Viewing Sequence Data

An analysis can be started by **first** viewing information about specific sequence data:

1. On the Home Page, select **Data Management**.
2. Click **View or Import Sequence Data**.
3. Click **Raw Data, BAM** or **Raw Data, RS II** to specify what type of sequence data to view. (To **narrow** the list of sequence data, enter the first few characters of the desired Data Set name in the Search field.)



4. In the **Name** column, click the name of the sequence data of interest. Details for the selected sequence data display.



5. To start an analysis using this sequence data, click **Analyze**, then follow the instructions starting at Step 5 of "Creating and Starting Analyses" on page 15.

Canceling a Running Analysis

Note: You can only cancel an analysis while it is currently running, **not** when it is in a Queued state.

1. On the Home Page, select **SMRT Analysis**.
2. Click the **Running** button to see only running analyses.
3. Select a currently-running analysis to cancel.
4. Click **Cancel**.
5. Click **Yes** in the confirmation dialog.

PacBio® Secondary Analysis Applications

Following are the secondary analysis applications provided with SMRT Analysis v4.0.0. For more details on the individual applications, see the **SMRT Link Online Help** by clicking the **?** icon in the upper-left corner of the screen.

Assembly (HGAP 4)

- Generate *de novo* assemblies of genomes.

Barcoding

- Analyze multiplexed samples.

Base Modification Detection

- Identify putative sites of base modification as well as common bacterial base modifications (6-mA, 4-mC, and optionally TET-converted 5-mC).

Base Modification and Motif Analysis

- Identify putative sites of base modification as well as common bacterial base modifications (6-mA, 4-mC, and optionally TET-converted 5-mC), and then analyze the methyltransferase recognition motifs.

CCS Mapping

- Generate consensus sequences from single molecules, and map these consensus sequences to a user-provided reference sequence.

CCS with Barcoding

- Identify consensus sequences for single molecules, for multiplexed samples.

Circular Consensus Sequences (CCS 2)

- Identify consensus sequences for single molecules.

Convert BAM to FASTX

- Convert sequence data in BAM file format to the FASTX file format.
- Currently, you must **first** run the **Barcoding** application to create BAM files **before** using this application.

Convert RS to BAM

- Convert sequence data generated on a PacBio RS II system in HDF5 file format to the BAM file format, compatible with SMRT Analysis v4.0.0.

Iso-Seq™ Algorithm

- Characterize transcripts and splice variants.

Iso-Seq Algorithm Classify Only

- This is only the **Classify** step of the Iso-Seq™ algorithm. Sequencing reads are classified into full length or non-full length reads.

Iso-Seq Algorithm with Mapping

- Characterize transcripts and splice variants, then map the transcripts back to the reference genome.

Long Amplicon Analysis (LAA 2)

- Identify phased consensus sequences from a heterogeneous pool of amplicons.

Long Amplicon Analysis with Barcoding

- Identify phased consensus sequences from a heterogeneous pool of amplicons, using multiplexed samples.

Resequencing

- Map sequencing reads against a reference sequence and identify variants.

Site Acceptance Test (SAT)

- Generate a report displaying instrument acceptance test metrics. (The application is designed **only** for analysis of Site Acceptance data.)

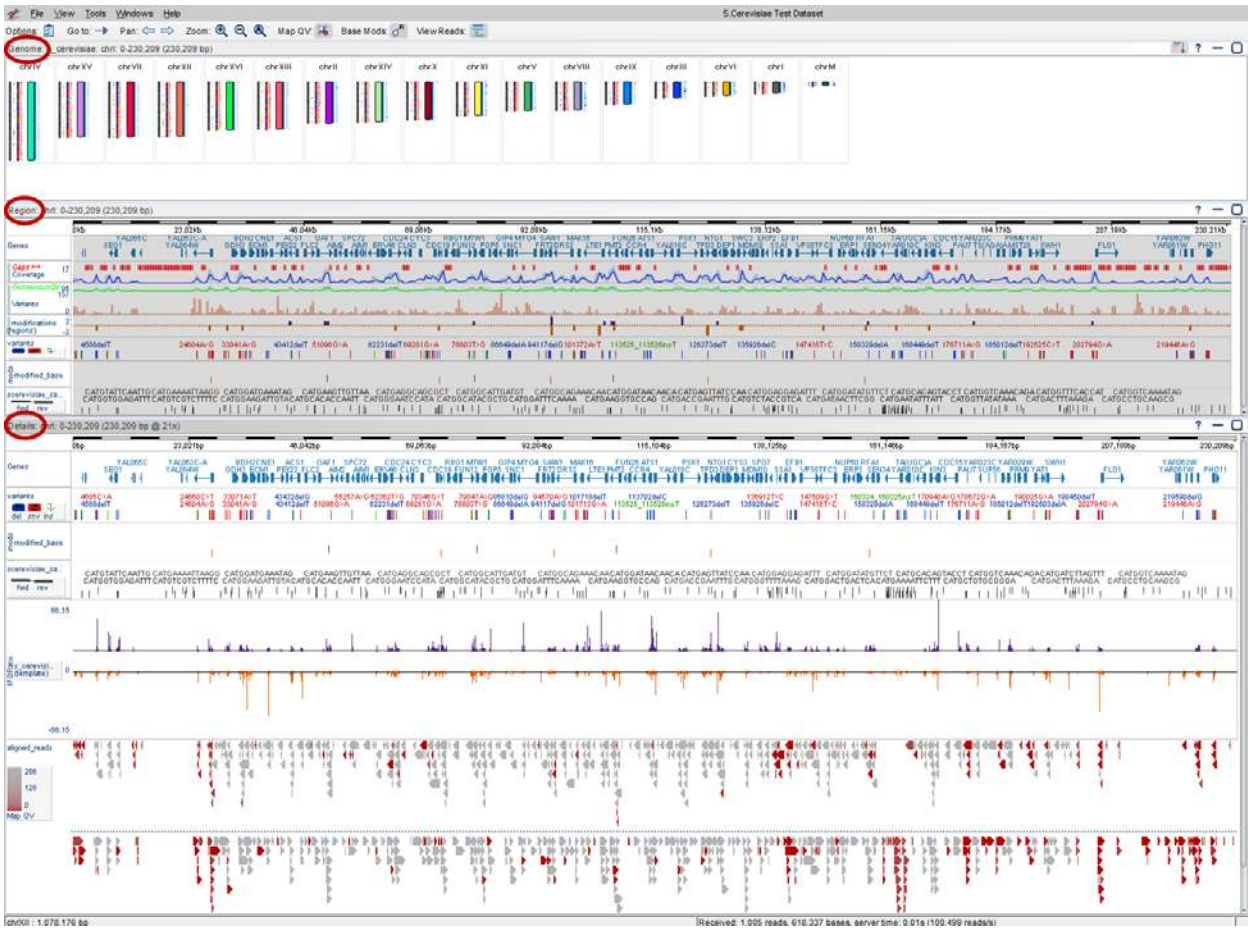
Visualizing Data Using SMRT® View

Once an analysis has successfully completed, visualize the results using **SMRT View**; a genome browser that displays sequencing data generated by the Sequel System.

Using SMRT View

1. In SMRT Link, select **SMRT Analysis**. A list of **all** analyses displays.
2. (Optional) Click the **Successful** button to see only successfully-completed analyses.
3. Click the name of a successfully-completed analysis to visualize. (**Note**: The analysis **must** have produced alignments as output for the SMRT View button to display.)
4. Click the **SMRT View** button located at the upper-right of the page.

5. The SMRT View application downloads to your computer and displays the data in three panels: **Genome**, **Region**, and **Details**.



- The **Genome** panel displays whole chromosomes or DNA segments, along with significant points of interest. The panel displays **only** if the secondary analysis data includes multiple genomes, chromosomes, or segments.
- The **Region** panel acts as a summary of the data. It displays metrics such as coverage and variants, and allows fast navigation across data to identify regions of interest.
- The **Details** panel allows drilling down to base-level resolution and visualizing SNPs, indels and kinetics used for base modification detection.

6. (Optional) **Click** a genome, chromosome or DNA segment to select it. Or, click and drag to select a section of interest. The Region panel displays the selection in greater detail.
7. In the **Region** panel, click and drag to select a smaller section. That section displays in the Details panel.
8. In the **Details** panel, click and drag to view the smallest area, down to the individual bases.

Note: To run SMRT View, 64-bit Java (Version 8 or later) **must** be installed on your local Windows or Macintosh computer.

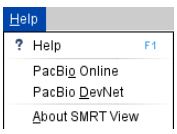
Installing 64-bit Java 8 on a Windows operating system

1. Use **Control Panel > Programs and Features** to check for and uninstall **all** existing versions of Java software.
2. Go to http://www.java.com/en/download/manual.jsp.
3. Click **Windows Offline (64-bit)**. This downloads a `x64.exe` file. (**Note**: Other Java versions are 32-bit, and will **not** work with SMRT View).
4. Double-click the `.exe` file to start the Java installer, and follow the installer directions.
5. After the installation is finished, restart the browser.

Installing 64-bit Java 8 on Mac OS (Note: This requires OS X 10.7.3 or later.)

1. Use the Finder to search for **all** existing versions of Java software, then drag them to the Trash to uninstall.
2. Go to http://www.java.com/en/download/manual.jsp.
3. Click **Mac OS X**. This downloads a `x64.dmg` file.
4. Double-click the `.dmg` file to mount the installer volume.
5. Double-click the Java icon to start the Java installer, and follow the installer directions.
6. After the installation is finished, restart the browser.

For more information on visualizing data, see the **SMRT® View Online Help**. To access the help, choose **Help > Help**.

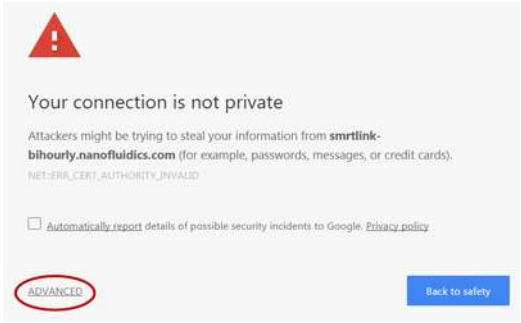


If Using the PacBio Self-Signed SSL Certificate

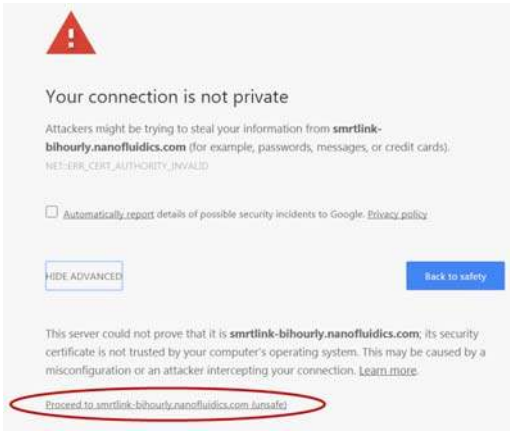
SMRT Link v4.0.0 ships with a PacBio self-signed SSL Certificate. If this is used at your site, security messages display when you try to login to SMRT Link for the **first time** using the Chrome browser. These messages may also display **other times** when accessing SMRT Link.

1. The first time you start SMRT Link after installation, you see the following text. Click **Open**.

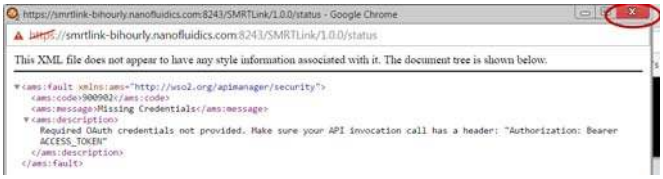
2. Click the **Advanced** link.



3. Click the **Proceed...** link. (You may need to scroll down.)



4. Close the window by clicking the **Close** box in the corner.



5. The **Login** dialog displays, where you enter the User Name and Password. The next time you access SMRT Link, the Login dialog displays **directly**.

Sequel™ System Output Files

This section describes the data generated by the PacBio® Sequel System for each cell that are transferred to network storage.

File Structure

Following is a sample of the file and directory structure output by the Sequel System:

```
<your_specified_output_directory>/r54008_20160116_003347/1_A01
|-- m54008_160116_003634.scraps.bam
|-- m54008_160116_003634.scraps.bam.pbi
|-- m54008_160116_003634.subreads.bam
|-- m54008_160116_003634.subreads.bam.pbi
|-- m54008_160116_003634.subreadset.xml
|-- m54008_160116_003634.sts.xml
|-- m54008_160116_003634.control.scraps.bam
|-- m54008_160116_003634.control.scraps.bam.pbi
|-- m54008_160116_003634.control.subreads.bam
|-- m54008_160116_003634.control.subreads.bam.pbi
|-- m54008_160116_003634.transferdone
```

In this example, `/r54008_20160116_003347` is a directory containing the output files associated with **one** run.

  - `r54008` is the instrument ID number
  - `20160116_003347` is the run **date**, in `YYYYMMDD` format, and **time**, in UTC format.
- The run directory includes a subdirectory for **each** collection/cell associated with a sample well-in this case `1_A01`. The collection/cell subdirectory contains output files of interest, described in this document.

Subreads.BAM File

The Sequel System outputs **one** `subreads.bam` file per collection/cell, which contains unaligned base calls from high-quality regions. This file is transferred from the instrument to network storage, then is used as **input** for secondary analysis by Pacific Biosciences' SMRT® Analysis software.

Data in a `subreads.bam` file is analysis-ready; all of the data present should be quality-filtered for downstream analyses. Subreads that contain information such as double-adapter inserts or single-molecule artifacts are **not** used in secondary analysis, and are excluded from this file and placed in `scraps.bam`.

- The BAM format is a binary, compressed, record-oriented container format for raw or aligned sequence reads. The associated SAM format is a text representation of the same data. The BAM specifications are maintained by the SAM/BAM Format Specification Working Group.

- BAM files produced by the Sequel System are **fully compatible** with the BAM specification.
  For more information on the BAM file format specifications, see **http://pacbiofileformats.readthedocs.io/en/3.0/BAM.html**.

BAM.PBI File

Pacific Biosciences' previous alignment file format (`cmp.h5`) contained a data table (the **alignment index**) that recorded auxiliary identifying information and precomputed summary statistics per aligned read. This table:

- Enabled fast random access to aligned reads satisfying fairly complex searches, for example, reads from a specific list of ZMWs which had unambiguous mapping (MapQV==254), or a read with a given read name.
- Allowed summary reports (read length, mapped identity/accuracy, and so on) to be constructed by quick operations over the alignment index instead of loading all of the sequence reads for each analysis.

To provide backwards-compatibility with the APIs enabled for accessing the `cmp.h5` file, a new BAM companion file was created - the **PacBio BAM index**, which supports the two use cases above.

For more information on the Pacific Biosciences BAM.PBI file format specifications, see **http://pacbiofileformats.readthedocs.io/en/3.0PacBioBamIndex.html**.

Other Output Files

- `.subreadset.xml`**:** This file is needed to import data into SMRT Link.
- `.scraps.bam` and `.scraps.bam.pbi`**:** These files contain sequence data outside of the High Quality region, rejected subreads, excised adapter and possible barcode sequences, as well as spike-in control sequences. (The basecaller marks regions of single molecule sequence activity as high-quality.)

  Note: This applies to files generated by Sequel™ Instrument Control Software (ICS) v3.1.0 or later.
- `.sts.xml`**:** Contains summary statistics about the collection/cell and its post-processing.
- `.control.`files**:** Contains sequence data for spike-in-control reads.
- `.transferdone`**:** Contains a list of files successfully transferred.

Frequently Asked Questions

What are the minimum files needed to analyze data on SMRT Link?

- `.bam` file
- `bam.pbi` file
- `subreadset.xml` file

What is the average size of the file bundle for a 6-hour movie?

Approximately 5 Gb.

What is the difference between a regular .bam file and an aligned.bam file?

The `subreads.bam` file contains all the subreads sequences, while the `aligned.bam` file additionally contains the genomic coordinates of the reads mapped to a reference sequence.

The `subreads.bam` file is created by the PacBio Sequel system, while the `aligned.bam` file is created by SMRT Link after running Resequencing or Mapping analysis applications.

LDAP

SMRT Link supports the use of LDAP for user login and authentication. **Without** LDAP enabled, only **one** common user is enabled, with **no** authentication or logging in required. SMRT Link **must** be integrated and configured to work with LDAP at your site **before** you can add SMRT Link users, or modify their roles.

- For details on integrating LDAP and SMRT Link, see the document **SMRT Link Software Installation (v4.0.0)**.

SSL

SMRT Link allows the use of Secure Sockets Layer (SSL) to enable access via HTTP over SSL (HTTPS), so that SMRT Link logins and data are encrypted during transport to and from SMRT Link. SMRT Link includes an Identity Server, which can be configured to integrate with your LDAP/AD servers and enable user authentication using your organizations' user name and password. To ensure a secure connection between the SMRT Link server and your browser, the SSL Certificate can be installed **after** completing SMRT Link installation.

It is important to note that PacBio will **not** provide a Signed SSL Certificate, however – once your site has obtained one – PacBio tools can be used to install it and configure SMRT Link to use it. You will need a certificate issued by a Certificate Authority (CA, sometimes referred to as a 'certification authority'). PacBio has tested SMRT Link with certificates from the following certificate vendors: VeriSign, Thawte and digicert.

Note: Pacific Biosciences recommends that you consult your IT administrator about obtaining an SSL Certificate.

Alternatively, you can use your site's self-signed Certificate.

SMRT Link ships with a PacBio self-signed SSL Certificate. If used, **each** user will need to accept the browser warnings related to access in an insecure environment. Otherwise, your IT administrator can configure desktops to always trust the provided self-signed Certificate. Note that SMRT Link is installed within your organization's secure network, behind your organization's firewall.

- For details on updating SMRT Link to use an SSL Certificate, see the document **SMRT Link Software Installation (v4.0.0)**.

The following procedures are available **only** for SMRT Link users whose role is **Admin**.

Adding and Deleting SMRT Link Users

1. Choose **Configure > User Management**.
2. There are 2 ways to find users:

- To display **all** SMRT Link users: Click **Display all Enabled Users**.

- To find a specific user: Enter a user name, or partial name, and click **Search By Name**.

3. Click the desired user. If the Status is **Enabled**, the user has access to SMRT Link; **Disabled** mean the user **cannot** access SMRT Link.
   - To **add** a SMRT Link user: Click the **Enabled** button, then assign a role. (See below for details.)
   - To **delete** a SMRT Link user: Click the **Disabled** button.
4. Click **Save**.

Assigning User Roles

SMRT Link supports three user roles: **Admin**, **Lab Tech**, and **Bioinformatician**. Roles define which SMRT Link modules a user can access. The following table lists the privileges associated with the three user roles:

Tasks/Privileges	Admin	Lab Tech	Bioinformatician
Add/Delete SMRT Link Users	Y	N	N
Access Run Design Module	Y	Y	N
Access SMRT Analysis Module	Y	Y	Y
Access Run QC Module	Y	Y	Y
Access Data Management Module	Y	Y	Y

1. Choose **Configure > User Management**.
2. There are 2 ways to find users:
   - To display **all** SMRT Link users: Click **Display all Enabled Users**.
   - To find a specific user: Enter a user name, or partial name, and click **Search By Name**.
3. Click the desired user.
4. Click the **Role** field and select one of the three roles. (A **blank** role means that this user **cannot** access SMRT Link.)
   - **Note**: There can be **multiple** users with the Admin role; but there **must** always be at least **one** Admin user.
5. Click **Save**.

P/N 100-976-100-01