

Guide - Step-By-Step Run Performance Evaluation

Introduction

This guide provides information on how to troubleshoot sub-optimal performance of PacBio® SMRT® sequencing runs using the Internal Control and primary metrics immediately available upon run completion. It is intended to help customers understand and interpret the metrics most important to raw sequencing performance and downstream secondary analysis applications.

Step 1: Evaluate the Performance of the DNA Internal Control

What is the Sequel DNA Internal Control Complex and Why Use It?

The DNA Internal Control Complex is a previously prepared bound complex used as a spike-in sequencing control on Sequel® Systems. It is composed of a fixed insert of 1966 bps with ligated SMRTbell® adapters, annealed primer and a bound polymerase. This control complex is spiked into the bound sample at the end of SMRT® Link Sample Setup and is intended to be a known ideal sample for monitoring the sequencing performance of Sequel Systems. Poor or unexpected performance of this control could indicate potential issues with the instrument, sequencing reagents, or consumables.

The sequence of the control does not have identity to any known organism and therefore can be easily detected and separated from sample data (see Additional Information for sequence). Control read filtering occurs on the instrument and all detected control reads are sent to the scraps.bam file. A few low-quality control reads may pass through the filter to the subreads.bam file, but this leak rate should be rare (less than 1% of the total number of control reads from a given SMRT® Cell).

How to Prepare the DNA Internal Control Complex for Sequencing

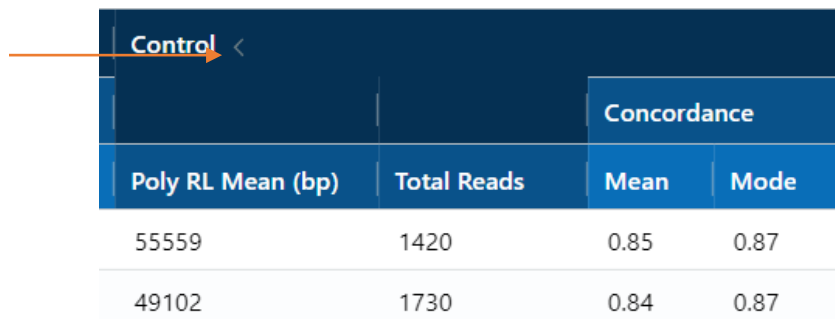
The DNA Internal Control Complex is provided as a stock solution and requires a serial dilution prior to adding it directly to the sample. It is critical to appropriately follow the directions for preparing the control complex as indicated in SMRT Link Sample Setup or the application-specific Procedure & Checklist.

The Control Complex should be kept on ice when handling and, after use, promptly returned to storage at -20°C to maintain stability. The stock tube, and all subsequent dilutions, must be fully homogenized by gentle finger tapping and spinning down before use. Do not reuse preparations of the control from previous runs. Typical Control Complex read counts are shown in the “Sequel Systems Control Concordance Too Low?” section (Table 1 for SMRT Cell 1M and Table 2 for SMRT Cell 8M).

How to use the DNA Internal Control Sequencing Performance Metrics for Troubleshooting

When a sequencing performance issue is observed for a sample, users should first evaluate the primary sequencing metrics of the DNA Internal Control. The Run QC table (Figure 1) in SMRT Link contains primary metrics for the control.

Click on the small arrow (<) next to "Control" to see full table.



Control <		Concordance	
Poly RL Mean (bp)	Total Reads	Mean	Mode
55559	1420	0.85	0.87
49102	1730	0.84	0.87

Figure 1: DNA Internal Control sequencing metrics table reported in SMRT Link Run QC

The expected value for each metric can be found in the table referenced in the relevant bullet point. The expected values are averages and reflect typical expectations from a correctly prepared control with the noted Run Design conditions.

- **Control Total Reads:** the total number of control reads detected (see Table 1 and Table 2 in the “Sequel Systems Control Concordance Too Low?” section).
- **Control Poly RL Mean:** the average polymerase read length of the control, including insert, adapters, and all passes around the SMRTbell template (see Table 3 and Table 4 in the “Sequel Systems Control Concordance Too Low?” section).
- **Control Concordance Mean/Mode:** the percent agreement of the control raw reads to the control reference sequence (see Table 5 in the “Sequel Systems Control Concordance Too Low?” section).

The Control Poly RL Mean and Control Total Reads of the DNA Internal Control Complex are common metrics used to assess sequencing run performance. To successfully interpret these metrics the Run Design parameters of the acquisition (i.e. movie length, pre-extension time, etc.) also need to be considered.

Generally, the Control Total Reads metric negatively correlates with increasing pre-extension time due to the normal loss of active complexes during incubation. The Control Total Reads metric also negatively correlates with increasing on-plate loading concentration (OPLC) of the sample due to competition between molecules for binding sites within the ZMWs of a SMRT Cell.

Additionally, increasing OPLC can result in poor Control Poly RL Mean as the increased number of complexes can cause ZMWs to be doubly occupied which results in a late start to high quality sequencing regions (see Step 3 and Figure 6). Also, the movie time greatly affects the Control Poly RL Mean. Given all these factors and trade-offs, consider the following points when interpreting the DNA Internal Control metrics and following the Troubleshooting Flow Chart (Figure 2):

- The Control Total Reads is typically >500. See Table 1 and Table 2 for typical counts for various Run Design parameters.
- The Control Poly RL Mean is typically >15 kb for 15-hour collections and >30 kb for 30-hour collection times. See Table 3 and Table 4 for typical Control read lengths for various Run Design parameters.
- If the Control Total Reads were <500, and no operator errors during preparation of the internal control are suspected, then troubleshooting should be directed at the PacBio system and consumables. In very rare cases, low control read counts can result from a contaminant in the sample that interfered with immobilization of the control complex on a SMRT Cell. Another possibility to consider is that using an excessively high OPLC can also lead to lowered control read counts.
- If the Control Total Reads were >500 and the Control Poly RL Mean were much lower than expected, then troubleshooting should be directed at the PacBio system and consumables. Low control read lengths can commonly indicate a nitrogen gas supply issue. Check the nitrogen supply source and pressure readings with your facilities team (Sequel II systems are typically ~50 PSI – however, your Pacific Biosciences Site Preparation Guide should be consulted for the most up to date recommendations). Other common causes of shortened control read counts include failure of the oxygen scavenging enzyme due to improper storage of the Sequencing Kit plate or using too high of an OPLC for the sample. In rare cases, a contaminant in the sample can interfere with polymerization of the control complexes.
- If both the Control Total Reads and Control Poly RL Mean performed as expected, then troubleshooting should be directed at the sample because the control performed within expectations and the instrument and consumables are unlikely to be the root cause.

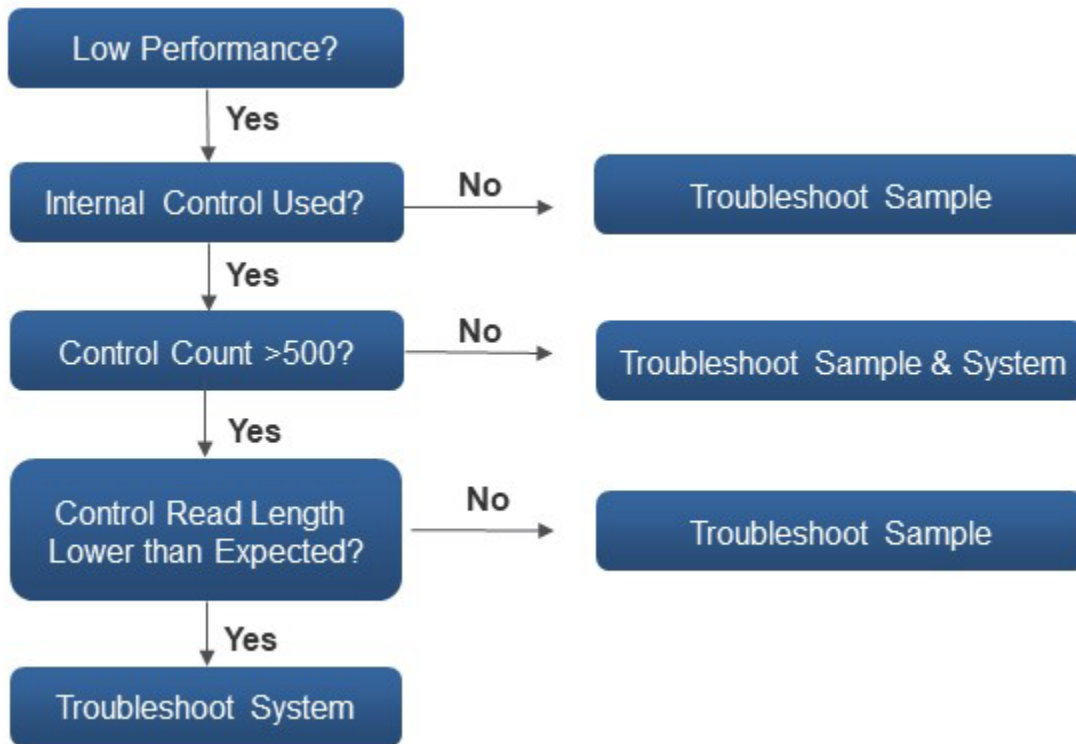


Figure 2: Flow chart to help guide Control metric troubleshooting

Sequel Systems Control Read Counts Too High?

An excessively high control read count can be due to out of calibration pipettors or deviation from Sample Setup instructions (1:10,000 dilution). These should be checked first. Avoid pipetting volumes $<1 \mu\text{L}$ to minimize pipetting errors. If control reads are consistently observed to be above $>20\text{k}$ reads, contact PacBio Technical Support for assistance. An additional dilution step may be performed to reduce the number of reads. Start with an additional 1:1 dilution after performing the initial control stock solution as per SMRT Link Sample Setup instructions.

Sequel Systems Control Concordance Too Low?

The expected values for the Control Concordance Mean are shown in Table 5 for the current Sequel® Systems and chemistry versions. Generally, over-loaded SMRT Cells will have a lower concordance score (<0.84), whereas optimally loaded SMRT Cells will have slightly higher concordance scores >0.84.

If the Control Concordance Mean is lower than expected, assess the P0 productivity metric to determine if the sample could have been over-loaded. Consider re-running the sample at a lower OPLC if the P0 productivity metric is <10%. Poor sample and consumable quality, or instrument issues could also affect the Control concordance, but these are not common.

Pre-Extension Time (mins)	Sequel System Plate 3.0	
	Median (count)	Quartile Range, 25%-75% (count)
240	6,900	4,000 - 10,200
120	12,200	6,100 - 19,800
0	13,000	7,100 - 21,800

Table 1: Average Control Read Counts for the Sequel System. Run conditions: SMRT Cells 1M v3 with Sequel DNA Internal Control Complex 3.0, Sequel Sequencing Kit 3.0 on ICS v8.0 software. (Values presented as median from field data).

Pre-Extension Time (mins)	Sequel II System Plate 2.0	
	Median (count)	Quartile Range, 25%-75% (count)
240	3,300	2,200 - 4600
120	5,500	3,400 - 8400
0	5,700	3,000 - 10,000

Table 2: Average Control Read Counts for the Sequel II System. Run conditions: SMRT Cells 8M with Sequel II DNA Internal Control Complex 1.0, Sequel II Sequencing Kit 2.0 on ICS v8.0 software. (Values presented as median from field data).

Movie Length (mins)	Pre-Extension Time (mins)					
	0		120		240	
	Median (kb)	Quartile Range, 25%-75% (kb)	Median (kb)	Quartile Range, 25%-75% (kb)	Median (kb)	Quartile Range, 25%-75% (kb)
1200	46,200	41,300 – 49,300	50,300	45,300 – 54,000	46,900	41,500 – 52,500
600	35,200	31,900 – 37,600	37,500	33,900 – 41,900	37,900	33,900 – 43,400

Table 3: Average Control read length for the Sequel System. Run conditions: SMRT Cells 1M v3 with Sequel DNA Internal Control Complex 3.0, and Sequel Sequencing Kit 3.0 on ICS v8.0 software.

Movie Length (mins)	Pre-Extension Time (mins)					
	0		120		240	
	Median (kb)	Quartile Range, 25%-75% (kb)	Median (kb)	Quartile Range, 25%-75% (kb)	Median (kb)	Quartile Range, 25%-75% (kb)
1800	56,000	47,400 -61,500	61,400	57,700 -65,300	58,100	53,500 - 62,100
1440	--	--	58,000	55,500 - 60,800	--	--
900	37,000	34,100 -39,000	--	--	--	--

Table 4: Average Control read length for the Sequel II System. Run conditions: SMRT Cells 8M with Sequel II DNA Internal Control Complex 1.0, Sequel II Sequencing 2.0 Kit on ICS v8.0 software.

System	SQ Plate 3.0		SQII Plate 2.0	
	Median (concordance)	Quartile Range, 25%-75% (concordance)	Median (concordance)	Quartile Range, 25%-75% (concordance)
Sequel	0.862	0.857 – 0.867	--	--
Sequel II	--	--	0.849	0.845 – 0.854

Table 5: Average Control Mean Concordance for Sequel and Sequel II Systems for raw data. Run conditions: SMRT Cells 1M v3 with Sequel DNA Internal Control Complex 3.0; or SMRT Cells 8M with Sequel II DNA Internal Control Complex 1.0; on ICS v8.0 software. (Values presented as median from field data).

Step 2: Verify SMRT Link Sample Setup Parameters are Correctly Defined

The SMRT Link Sample Setup section contains a list of all sample sheets generated and saved by the user(s). In cases where poor sequencing performance was obtained for a sample, verify that the Sample Setup sheet was correctly generated for the specific sample in question. Common mistakes can be identified and rectified if the same sample is to be run again. The user should:

- Ensure that the latest version of SMRT Link is installed. Detailed information on how to check the version is found in the [Sequel II System Operations Guide](#). The latest version is always listed [here](#). If you do not have the latest version, check with your IT department and PacBio about upgrading your software.
- Check that the correct binding kit was used for your application and insert size. Using the incorrect binding kit can cause low loading and low total base yields. More information about binding kit recommendations can be found in: [Overview – Sequel Systems Application Options and Sequencing](#)
- Pay close attention to any warning messages that may appear in Sample Setup. For example, if Sample Setup warns that the sample is too dilute, consider concentrating the sample with AMPure® PB/ProNex beads or, if possible, prepare additional SMRTbell libraries.
- As the polymerase dilution is variable, pay close attention to any extra dilutions that may be required. The polymerase should be diluted fresh for each run, kept on ice while in use, and diluted polymerase should not be stored.
- Double-check that the entry in the sample Concentration field matches the Qubit assay measurement. PacBio strongly recommends using the Qubit dsDNA HS kit for accurate quantification of double-stranded DNA and final SMRTbell libraries.
- Check that the entry in the Insert Size field matches the sizing QC measurement for your library's insert size. For libraries >15 kb, PacBio recommends using a pulsed-field electrophoresis method. For assessing shorter libraries, the Agilent Bioanalyzer system is a good choice.
- The Concentration and Insert Size are used to automatically calculate the molarity of the sample (Figure 3). Accurate molarity determination is critical for generating optimized reaction conditions for primer annealing, polymerase binding and sample immobilization on the SMRT Cell.

$$\frac{\text{Insert Concentration (ng/}\mu\text{L)}}{\text{Mean insert size (bp)}} * \frac{1 \text{ mol basepairs}}{650 \text{ g}} * \frac{1 \text{ g}}{10^9 \text{ ng}} * \frac{10^{15} \text{ fmol}}{1 \text{ mol base pairs}} = \text{fmol} = \text{nM}$$

Figure 3: SMRTbell Molarity Calculation used in SMRT Link v8.0 Sample Setup.

- Check the sample AMPure PB bead/ProNex bead Cleanup Yield % after performing polymerase-bound complex cleanup. The typical % recoveries for complexes with various insert sizes are shown in Table 6. In general, the efficiency of recovery is lower for complexes with shorter insert sizes. If the complex % recovery was significantly below expected values, then the sample may be degraded, and it may be best to start with a fresh sample.
- If requesting assistance from PacBio Technical Support, the Sample Setup sheet can easily be exported by using your browser's "Print" and "Save as PDF" function. To export PDF reports from SMRT Link, see the [SMRT Link User Guide](#)

Library Type and Insert Size	1.2X AMPure PB Bead Cleanup Recovery (%)
>30 kb large-insert libraries	>70
5 kb - 6 kb amplicon libraries	50-70
8 kb - 12 kb Microbial Multiplexing library	40-50
3 kb - 4 kb amplicon libraries	40-50
1 kb - 2 kb amplicon libraries	30-40
Iso-Seq libraries	30-50
250 bp - 1000 bp amplicon library	20-40

Table 6: Typical Recovery Cleanup Yield % for various insert sizes.

Step 3: Evaluate the Productivity Metrics

Productivity

The SMRT Link Run QC table (see Figure 1) includes primary metrics for productivity, commonly referred to as "loading" metrics (see below). The productivity metrics show how many ZMWs on the SMRT Cell have produced useful sequencing data. After completion of sequencing data acquisition for a given SMRT Cell, the Sequel Systems' primary analysis pipeline classifies each ZMW, and the sequencing read it produced, into one of the following three productivity categories:

- **P0:** An empty, non-productive ZMW with no signal detected.
- **P1:** A productive ZMW with a high quality (HQ) sequencing region detected within the read.
- **P2:** Other, signal detected but no HQ region was detected.

These metrics are represented as a percentage of the total number of ZMWs present on a SMRT Cell. The P0 metric indicates the number of ZMWs that have no detectable sequencing signal above baseline (i.e. pulsing from base incorporations) and are inferred to be empty of an active sequencing polymerase complex. A ZMW that produced ≥ 50 bases of high-quality sequencing (i.e. a HQ Region) at any time during acquisition will be classified as P1. These P1 reads are stored in the subreads.bam file and are used in downstream applications. All other ZMWs producing detectable signal, but not a HQ region, are classified as P2 and are filtered to the scraps.bam file. These ZMWs commonly have high and noisy background signal and may contain multiple active sequencing complexes per ZMW.

How to Use the Productivity Metrics to Evaluate and Troubleshoot Sequencing Performance

When evaluating the performance of a sequencing run, assess the loading performance of a SMRTbell complex by primarily considering the P0 metric. The P0 metric will generally provide the most accurate estimate of sample complex loading on the SMRT Cell. Adding excessive quantities of complex to a SMRT Cell can negatively affect the sequencing performance of the sample and the quality of downstream secondary analysis. As shown in Figure 4, once an optimal inflection point (here $\sim 20\%$ P0) of loading is reached, a higher degree of loading will result in poorer raw sequencing performance, as read lengths, yield and concordance metrics decline. Note that P1 reads still increased past this point, however the reads would be shorter and of lower raw accuracy. Also, it is important to note that the performance of the DNA Internal Control itself was negatively affected.

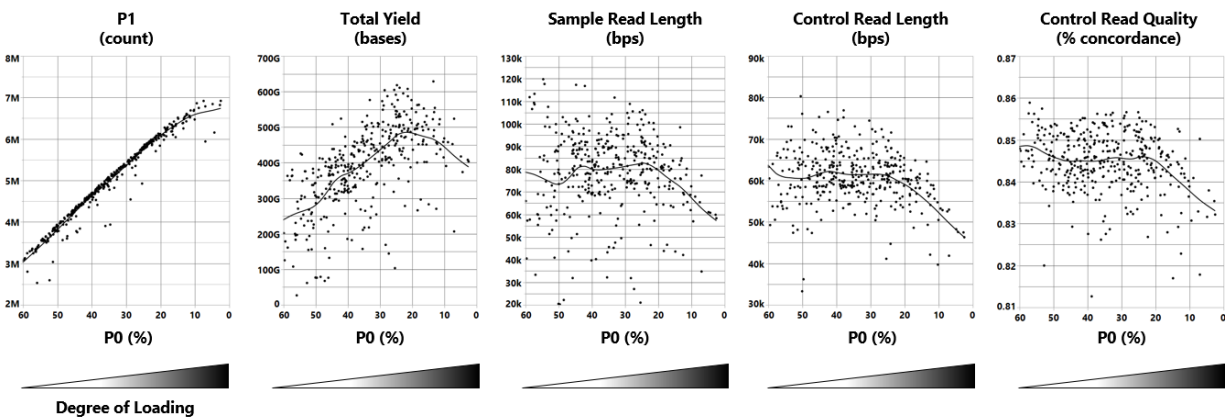


Figure 4: Raw performance metrics as a function of loading on Sequel II 8M SMRT Cells. Customer field data was analyzed after filtering for the following Run Design criteria: Insert size 10,000 to 20,000 bps, 30 hrs acquisition, 2 hrs pre-extension, SMRTbell Express Template Prep Kit 2.0, Sequel II Sequencing 2.0 Kit, Sequel II Binding Kit 2.0, and SMRT Link v8.0.

Conversely, a high P0 (>50%) would result in fewer P1 reads and lower Total Base yield, as fewer actively sequencing complexes would be present on the SMRT Cell. Mean polymerase read length may be slightly longer as fewer multi-loaded ZMWs would have been present under such conditions. See Figure 5 below.

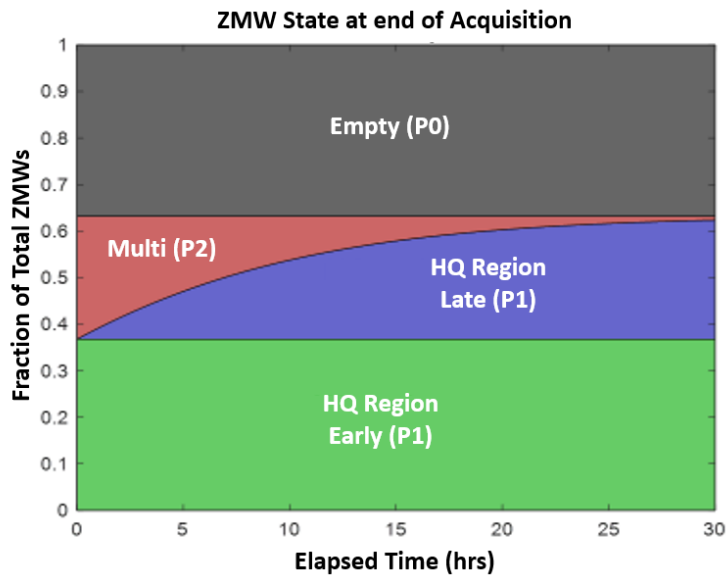


Figure 5. This figure shows a simulated time-course model of ZMW end-state productivity for a typical acquisition. The y-axis reflects the proportion of the total ZMWs in each state if the acquisition were to end at that moment (x-axis). Gray represents ZMWs that were always empty (“Empty (P0)”) throughout the acquisition and never produce a HQ sequencing read. Green indicates ZMWs that produced useful, high quality sequencing, immediately upon acquisition (“HQ Region Early (P1)”) and tend to produce the longest read lengths. Red indicates ZMWs that were loaded with two or more active polymerases (“Multi (P2)”) and at least two of these polymerases remained active. As acquisition progresses, active complexes in P2 ZMWs will terminate at a certain rate and only when one SMRTbell complex per ZMW is left sequencing will a HQ Region be detected and the ZMW state will turn to P1 (Blue, “HQ Region Late (P1)”). These ZMWs tend to produce the shortest read lengths.

Note that with higher loading on SMRT Cells, the proportion of ZMWs in the P2 state increases, and therefore, the number of ZMWs in the HQ Region Late state increase. The time taken for a P2 state ZMW to transition to a P1 state ZMW also lengthens. This results in far fewer P1 state ZMWs utilizing the full length of the acquisition time for useful, high quality, long polymerase read length sequencing.

Determining the optimal loading range will depend on balancing these interactions with the overall goal of the sequencing project. In applications where long polymerase read length is the priority, like long insert HiFi *de novo* assembly, then a SMRT Cell run with an optimal or slightly higher P0 may be more suitable.

For smaller insert applications where maximizing Total Base yield, P1 read count and sample throughput is the priority, and mean polymerase read length already exceeds the minimum requirements for secondary analysis, then a lower P0 may be tolerated. In general, for optimal raw sequencing performance, users should aim for 20% - 30% P0 for most applications.

Impact of SMRTbell Complex Over-Loading and Recommended Corrective Actions

Over-loading polymerase-bound complexes on a SMRT Cell will result in shorter Mean Polymerase Read Length of the sample and thus decrease the yield of Total Bases (Figure 4). As mentioned previously, for any given SMRT Cell there is typically a trade-off between the number of P1 reads generated and the Mean Polymerase Read Length.

For most types of long-read sequencing, a larger number of shorter reads may result in less useful data overall than generating a slightly smaller number of longer reads. A useful method for empirically determining optimal loading is to perform a loading titration of multiple OPLCs across multiple SMRT Cells from the same lot. With increasing OPLC, P0 will decrease, though not necessarily linearly. Once P0% decreases to <10% and mean Polymerase Read Length of the sample and the DNA internal Control start to decrease significantly, the sample concentration has most likely entered an over-loaded state. Alternative metrics of assessing whether a sample was over-loaded are generally not as reliable as the P0 metric. A high value of either P1 or P2 for a SMRT Cell, taken by itself, should not be automatically interpreted as over-loading. Instead, one should assess the P0 metric and polymerase read length.

If an individual SMRT Cell has shorter-than-expected Polymerase Read Length with lower-than-desired GB Yield, a P0 metric <10%, and sample quality is not in doubt, then it is generally recommended to re-run the sample at a lower OPLC. However, it is important to note that if there is any DNA damage, contamination, or other sample-related quality issues, then further troubleshooting may be necessary beyond adjusting the OPLC.

Impact of Complex Under-Loading and Recommended Corrective Actions

One of the most common, and easiest to fix, causes of low P1 reads and Total Base Yield is under-loading of polymerase-bound complexes onto the SMRT Cell. If the P0 metric for a SMRT Cell is substantially higher than 40%, P1 and P2 values are both low, and sample and control read lengths are as expected, then the SMRT Cell was likely under-loaded. To correct this, and obtain improved complex loading, the user can try re-running the sample at a higher OPLC. As complex loading on SMRT Cells is not a linear process, only relatively small increases in OPLC (1.2-2x) are usually required to see an impact on the productivity loading profile.

Another common cause of under-loading includes inefficient generation of polymerase-bound complexes due to operational mistakes during the primer annealing, polymerase binding or complex cleanup procedures (a.k.a. ABC). A useful indicator that can sometimes point to bound complex-related quality issues is shorter than expected polymerase read length on an under-loaded SMRT Cell. If there is any doubt regarding the generation of the bound complex then the user should first consider redoing ABC before attempting to re-run another SMRT Cell.

It is worthwhile to note that increasing OPLC will only help if there are no degradation, quality or contamination issues affecting the DNA sample. These issues may be detected by observing the polymerase read lengths for the sample. Typically, low quality samples will have severely reduced polymerase read length, show degradation on molecular weight sizing gels, and have low spectrophotometric absorbance ratios. Rerunning a "bad" sample at a higher OPLC will not rescue that sample. The quality of the sample should be evaluated, and likely re-extracted with additional purification steps and/or care to limit degradation and DNA damage.

Causes for High P2%

A high P2 loading profile can have several causes, although the most common is from over-loading the SMRT Cell. If you observe a SMRT Cell with a high P0, low P1 and high P2 loading profile (referred to as the "V-phenotype"), then there could be a contaminant present in the sample that is interfering with polymerase activity. The "V"-phenotype can also be caused by a drop in nitrogen pressure, or electrical issues within the SMRT Cell. In such instances, please contact PacBio Technical Support for troubleshooting assistance (support@pacb.com). Historically, high P2s were commonly caused by excess polymerase mistakenly added during complex binding (e.g. a dilution step was missed) or poor cleanup performance. AMPure PB cleanup is very effective at removing excess polymerase and this is rarely seen in practice.

Step 4: Evaluate the Insert Read Length Density (IRLD) Plot

IRLD Plot Features (see Figure 6):

- **Insert Read Length:** maximum (longest) subread length per ZMW.
- **HQ Read Length:** polymerase read length for a ZMW that produced ≥ 50 bases of high-quality sequencing.
- **Line A:** primary diagonal; terminations in first observed pass along insert (e.g. DNA damage, movie end), termination at 1st adapter; or polymerase fell off end of molecule (inefficient ligation).
- **Region B:** area between lines A and C; reads that terminated after the 1st observed adapter.
- **Line C:** secondary diagonal; terminations at or near 2nd observed adapter.
- **Region D:** primary CCS/HiFi generating region (if applicable).
- **Line E:** size selection boundary cut-off (if applicable).

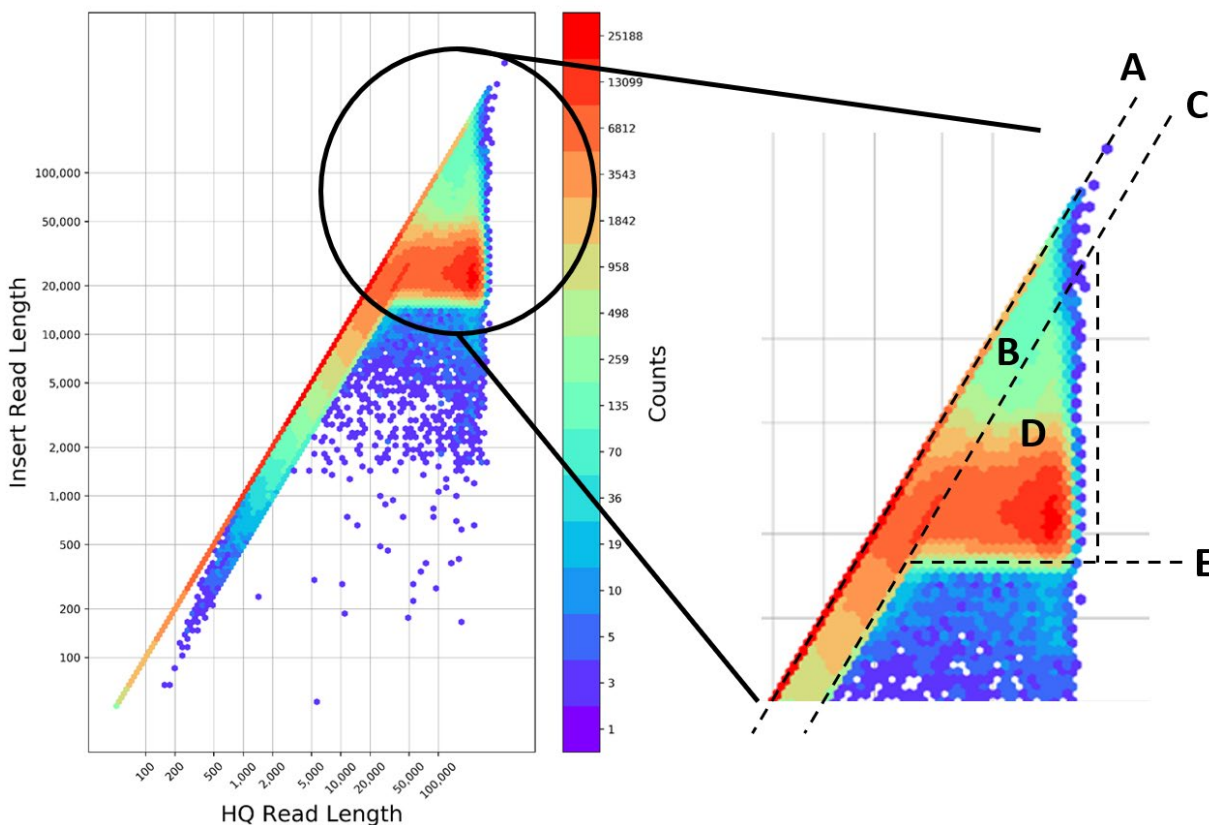


Figure 6: IRLD plot from a Sequel II System run showing useful troubleshooting features.

Description of the Plot

The IRLD plot, found in the Run QC section of SMRT Link, is a useful tool for visualizing and evaluating the distribution of SMRTbell insert sizes and polymerase read lengths in your sequencing dataset (see Figure 6). Each dot represents the amount of productive P1 ZMWs according to the HQ Read Length on the x-axis, and Insert Read Length on the y-axis (both axes are in log scale).

The HQ Read Length is the length of the HQ region of the sequencing read produced by the polymerase within a given ZMW and includes the insert and adapter portions as well as all passes around the template. The Insert Read Length represents the maximum subread length detected per ZMW and may indicate the size of the physical insert. It requires finding at least 2 adapter sequences to be estimated correctly. The density of reads is color coded, with purple representing low density of read counts, and red indicating high density of

reads. As the HQ Read Length is always \geq Insert Read Length, for a given ZMW, no data appears in the top-left region of the plot.

IRLD plots are extremely useful for interpreting basic characteristics of your SMRTbell library to aid in troubleshooting and can provide information such as: average insert size range, effectiveness of a library size selection step, presence of adapter ligation issues, polymerase early termination events, proportion of reads sequenced in CCS mode, etc.

The primary diagonal line (Line A) running in the center of the plot corresponds to reads where the HQ Read Length = Insert Read Length. Reads which fall along this line are reads in which no adapter was detected in the acquired HQ region. SMRTbell libraries made for CLR, with their very long insert sizes (>30 kb), will normally produce a high density of reads appearing along this diagonal. This is due to the inability of the polymerase to reach an adapter in very long insert SMRTbell libraries. Data from these ZMWs are exclusively generated from a single pass of the polymerase across the insert.

When the HQ Read Length $>$ Insert Read Length, the reads accumulate to the right of the primary diagonal and spread out along the x-axis. These are reads that are entering ≥ 2 observed passes along their insert, i.e. have seen ≥ 1 adapter.

A secondary diagonal line (Line C), shifted to the right and parallel to the primary diagonal, may also form and represents polymerase read lengths of 2x the SMRTbell insert size (as determined by adapter detection). This secondary diagonal represents termination of polymerases at or near the second observed adapter. This line usually represents a small proportion of the total reads in a dataset.

The data that falls between lines A and C (denoted as Line B), are those polymerases that sequenced passed one observed adapter and terminated sequencing before the second observed adapter. If a polymerase has sequenced >2 observed adapters the data point will fall to the right of the secondary diagonal (Line C) and indicates circular consensus sequencing (CCS)/rolling circle replication mode (region D).

Sample types, such as small amplicons, Iso-Seq and HiFi libraries, that are expected to enter CCS mode should show a large accumulation of data points in region D. In CCS mode, the polymerase is very stable and will produce the longest HQ Read Lengths. Reads in CCS terminate mostly because of laser-induced photodamage, or because the movie acquisition ended.

A third prominent line, typically found in size selected libraries, corresponds to the size selection cutoff boundary (Line E). It is denoted by a sharp drop in reads below a line parallel to the x-axis at the Insert Read Length of the size selection cutoff. If a strict size selection cutoff (e.g. Sage Science BluePippin system size selection) was utilized during sample preparation, then this boundary should be well defined. Line E represents the insert size at which molecules below that size were removed from the SMRTbell template population. If this line is not well defined, or is at the wrong insert size, then the size-selection step should be suspect.

Using the IRLD Plot to Evaluate Your Run

The pattern and interpretation of this plot is very much dependent on the application, how it was prepared, and the Run Design parameters used for data collection (e.g. pre-extension and acquisition length). For example, a CLR library used for *de novo* assembly, with a high size selection cutoff using a platform such as Sage Science BluePippin system and no pre-extension, will have a very different density distribution compared to a 2 kb PCR product. Additionally, the various size selection methods available will greatly influence this plot (see Figure 7). Interpretation of the IRLD plots is a valuable tool for troubleshooting sequencing runs and samples.

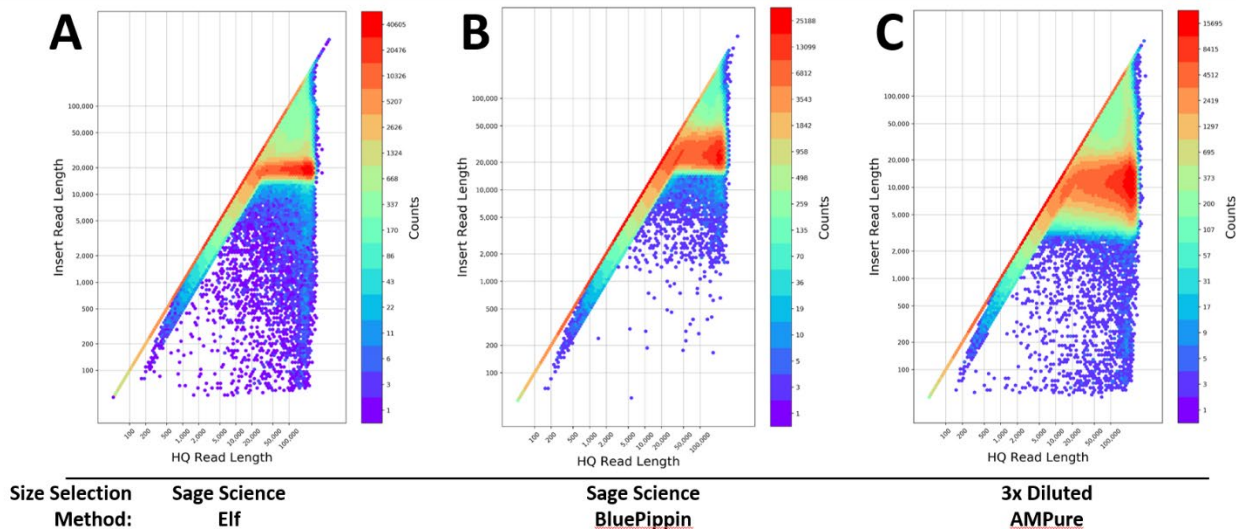


Figure 7: IRLD plots of HiFi SMRTbell libraries size selected by different methods and ran on Sequel II 8M SMRT Cells. A. Human library size selected using a Sage Elf system showing an ideal, relatively narrow, insert size range. B. Plant library size selected using a Sage BluePippin system showing a discrete size cutoff on the low end, but a broader range of larger fragments. C. Human library size selected using the 3x Diluted AMPure method showing the broadest and shortest insert size range. All datasets were acquired using 30 hrs acquisition, 2 hrs pre-extension, SMRTbell Express Template Prep Kit 2.0, Sequel II Sequencing 2.0 Kit, Sequel II Binding Kit 2.0, and SMRT Link v8.0

For large gDNA libraries, with inserts >30 kb, and without pre-extension, most of the reads should be above the target size-selection cutoff and fall on the primary diagonal. To a lesser extent, some reads would also be in regions B and D, i.e. some reads would be expected to contain 1 or more adapters and maybe even enter CCS mode. More reads are observed along the primary diagonal than in region D with increasing average insert size; increasing size-selection cutoff; increasing sample molecular weight; with decreasing movie acquisition time; or without pre-extension.

Seeing an unexpected high density of reads making it into CCS (region D) could indicate that the average insert size was much smaller than expected, that a small DNA insert contaminant was unsuccessfully removed, or was generated during ligation (e.g. adapter dimers), or size selection failed.

If a high density of reads is showing up lower down on the primary diagonal, and the HQ Read Length was shorter than expected, this would indicate early polymerase termination events. This is a common symptom with CLR and HiFi SMRTbell libraries that have DNA damage and/or polymerase inhibitors contaminating the sample. Additionally, inefficient ligation, overloaded SMRT Cells, too short of an acquisition time, or failure to use pre-extension can also be the cause, though they are not common. If the sample is suspected, then additional AMPure PB bead purifications of the starting gDNA or re-extraction via a different method are the best strategies for successful sequencing.

The IRLD plot is also useful for visualizing reads of short insert samples. These would be fixed size inserts such as amplicons (see Figure 8), or short, low molecular weight, non-size selected samples such as those used for microbial multiplexing or Iso-Seq analysis. All these sample types should generate a large abundance of CCS reads that appear to the right of the diagonals A and C (see Figure 6) where the HQ Read Length is greater than Insert Read Length, region D. The accumulation of reads shorter than your expected insert size and, usually shifted to the right, would indicate a population of shorter than expected inserts, for example short secondary PCR products or fragmented library.

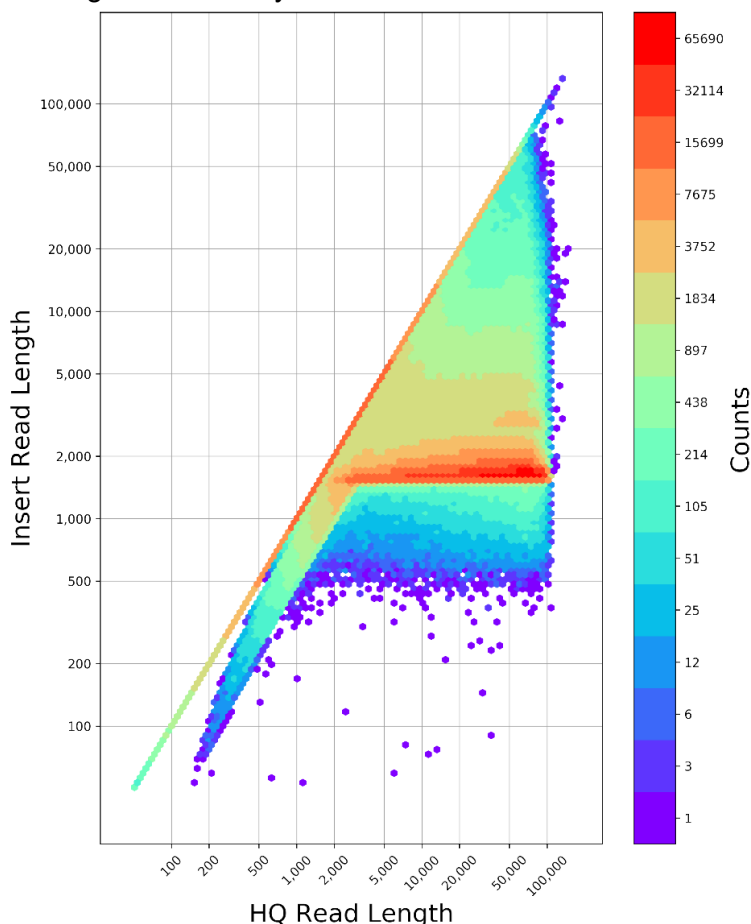


Figure 8: IRLD plot of a 1700 bp PCR amplicon SMRTbell library. A very discrete insert size was observed with a small accumulation of a chimera product at 2x the insert size. Dataset was acquired on a Sequel II 8M SMRT Cell using 10 hrs acquisition, 0.4 hrs pre-extension*, SMRTbell Express Template Prep Kit 2.0, Sequel II Sequencing 2.0 Kit, Sequel II Binding Kit 2.1, and SMRT Link v8.0.

For small, non-size selected libraries, the density of reads to the right of the diagonal lines A and C should closely reflect the expected size distribution of the library. If it does not, it could suggest quality issues with the library, or unequal molar pooling of samples. If you employed a size-selection step during sample preparation, then the lower boundary of the cutoff should be clearly distinguishable and at the expected insert size. If not, then the size selection step should be investigated.

Similar to large gDNA libraries, if a high density of reads are showing up lower down on the diagonal line A (see Figure 6), then that would indicate a reduced polymerase read length caused by DNA damage, polymerase inhibitors, or overloaded SMRT Cells. If there was an accumulation of reads higher up along the diagonal A, then that could indicate missing molecular adapters due to inefficient ligation. Similarly, the appearance and accumulation of reads at two times the expected insert size could indicate chimeras, which are end-to-end ligations of DNA products with adapters on both ends. A higher fraction of end-to-end ligated chimeric sequences are typically caused by adding too little adapter, or too much sample during ligation.

Note: The use of pre-extension can complicate the interpretation of the IRLD plot, because when acquisition starts, the polymerase may already be in CCS mode. Therefore, the number of physical passes is difficult to infer depending on the amount of pre-extension time used relative to the insert size of the sample. In general, pre-extension should increase the density of reads seen to the right of the diagonal line for inserts <15 kb.

Troubleshooting Guidance

Symptom	Potential Causes	Possible Actions/Solutions
Sample Shows Poor Sequencing Performance		
Sample shows short Polymerase Read Length (DNA internal control sequencing performance is normal)	Excessively high sample OPLC	Reduce sample OPLC
	Sample quality issue (e.g., highly fragmented DNA, high amount of DNA damage, presence of a contaminant)	Confirm sample QC; re-purify sample (consider changing methodology)
	Incorrect or insufficient pre-extension time specified in Run Design	Use SMRT Link recommended or longer pre-extension time setting.
	Inefficient adapter ligation reaction	Verify adapter ligation conditions used during SMRTbell library construction; reperform ligation reaction step if needed
Sample shows low P1 productivity metric (DNA internal control sequencing performance is normal)	Manual pipetting error during primer annealing, polymerase binding or complex cleanup (ABC) steps	Redo sample ABC steps
	Sample quality issue (e.g. highly fragmented DNA, high amount of DNA damage, presence of a contaminant)	Confirm sample QC (DNA molecular weight, purity and concentration); re-purify sample (consider changing methodology)
	Used pre-extension for a continuous long read sample	Rerun without pre-extension
High P2 productivity (DNA internal control sequencing performance is normal)	Excessively high sample OPLC	Reduce OPLC
	Sample quality issue (e.g. contaminant)	Confirm sample QC (DNA purity); re-purify sample (consider changing methodology)
	Excessively high unbound polymerase carryover	Redo polymerase binding step
	Use of non-recommended plastic consumables (leaching of fluorescent contaminants)	Use recommended plastic consumables for all sample extraction/purification, library construction and sequencing preparation steps
Insert Read Length Density plot shows larger than expected library insert size (DNA internal control sequencing performance is normal)	Concatemers present due to inefficient adapter ligation reaction	Verify adapter ligation conditions used during SMRTbell library construction; redo ligation reaction step if needed
	Sample quality issue (e.g. incorrect library insert size)	Confirm sample QC and molarity (DNA molecular weight); redo sample size selection step

Symptom	Potential Causes	Possible Actions/Solutions
DNA Internal Control Shows Poor Sequencing Performance		
Low Control Total Reads (Sample sequencing performance is either normal or poor)	Pipetting error during preparation of DNA Internal Control	Re-prepare DNA Internal Control using calibrated pipettes
	Sample quality issue (e.g. contaminant)	Confirm sample QC (DNA purity); re-purify sample (consider changing methodology)
	Excessively high sample on-plate loading (OPLC) concentration	Reduce sample OPLC
	SMRT Cell or instrument hardware issue	Contact PacBio Technical Support
Short Control Polymerase Read Length (Mean) (Sample sequencing performance is poor)	N2 interruption at source or on-instrument	Contact PacBio Technical Support
	Failure of biochemical O2 scavenging system due to use of expired or incorrectly stored sequencing plate	Use a new and correctly stored sequencing plate
	Sample quality issue (e.g. contaminant)	Confirm sample QC (DNA purity); re-purify sample (consider changing methodology)
	SMRT Cell or instrument hardware issue	Contact PacBio Technical Support
High Control Total Reads (Sample sequencing performance is either normal or poor)	Pipetting error during preparation of DNA Internal Control	Re-prepare DNA Internal Control using calibrated pipettes
	A dilution step was inadvertently omitted during preparation of the DNA Internal Control	Re-prepare DNA Internal Control
	Excessively low sample OPLC	Increase sample OPLC
	Instrument hardware issue	Contact PacBio Technical Support
Low Control Concordance (Sample sequencing performance is poor)	Excessively high sample OPLC	Reduce sample OPLC
	Sample quality issue (e.g. contaminant)	Confirm sample QC (DNA purity); re-purify sample (consider changing methodology)

Contacting PacBio for Support

If you suspect the consumables or the instrument are the root cause for the sub-optimal performance of your run, please reach out to your local FAS or to Technical Support for assistance. Additionally, if you are having difficulties with your sample, we may also be able to help. The Technical Support and Service teams in Menlo Park, California can be contacted at:

- Hours: Monday – Friday, 7:00 a.m. – 5:00 p.m. (PST)
- Email: support@pacb.com (preferred method of contact)
- Phone: 1.877.920.PACB (7222)

To speed along the support process, it is recommended to include as much relevant information as possible in your first email to us. Please include the following:

- SMRT Cell run information:
 - Exported PDF report from the Data Management section of SMRT Link
 - If not available: Run ID, movie ID, date, sample well, instrument serial number (64XXXX/54XXXX).
- Sample information:
 - Brief description of application (de novo assembly, Iso-Seq, etc)
 - What organism(s)/tissue(s) were the samples from?
 - Have you successfully sequenced from these organism(s)/tissue(s) before?
 - What extraction protocol was used?
 - What PacBio sample preparation procedure did you follow? (P/N and version)
 - What were the absorbance ratios of the starting material?
 - Was Qubit quantification performed on the starting material and did it correspond to the Nanodrop quantification?
 - Include any sizing gel images or traces performed, including RIN numbers if relevant.
 - Include any PDF reports of previous runs for the same sample, SMRTbell library or bound complex. To export PDF reports from SMRT Link, see the [SMRT Link User Guide](#).

Additional Information

- Resources for best practices during sample preparation, SMRTbell preparation protocols, and software documentation can all be found on www.pacb.com.
- Our extensive tutorial training slide decks contain useful information on all topics related to SMRT Sequencing, from sample extraction to data analysis: <https://www.pacb.com/support/training/>
- Sample preparation protocols for our template preparation kits for all applications: <https://www.pacb.com/support/documentation/>
- Helpful information on sample extraction best practices can be found: <https://www.pacb.com/wp-content/uploads/Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf> ...and here: <https://extractdnaforpacbio.com/>
- Useful table of experimental design, protocols, reagents, and Run Design conditions: [Overview – Sequel Systems Application Options and Sequencing](#)
- SMRT Link software user guides and release notes: <https://www.pacb.com/support/software-downloads/>
- Consideration for reimbursement of consumables lost due to sequencing run failures will be dependent on the use of the DNA Internal Control Complex. This reagent can help separate sample from system causes on failed runs with no obvious hardware or software issues.
- The sequence of the Sequel DNA Internal Control Complex (1996 bps) can be found on NCBI, accession MG495226: <https://www.ncbi.nlm.nih.gov/nucleotide/MG495226>. The unrolled sequence (including both strands plus adapters, 4034 bps) can also be found on NCBI, accession MG551957.1: <https://www.ncbi.nlm.nih.gov/nucleotide/MG551957.1>.
- [Pacific Biosciences Glossary of Terms](#)

Revision History (Description)	Version	Date
Initial Release	01	August 2020

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020, 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 to the applicable license terms at <http://www.pacificbiosciences.com/licenses.html>. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq and Sequel are trademarks of Pacific Biosciences. Femto Pulse and Fragment Analyzer are trademarks of Agilent Technologies, Inc. All other trademarks are the sole property of their respective owners.