

Research use only. Not for use in diagnostic procedures.

P/N 103-380-300 Version 01 (April 2024)

© 2024 Pacific Biosciences of California, Inc. ("PacBio")

Information in this document is subject to change without notice. PacBio assumes no responsibility for any errors or omissions in this document.

PACBIO DISCLAIMS ALL WARRANTIES WITH RESPECT TO THIS DOCUMENT, EXPRESS, STATUTORY, IMPLIED OR OTHERWISE, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY, SATISFACTORY QUALITY, NONINFRINGEMENT OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL PACBIO BE LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY, PURSUANT TO ANY STATUTE, OR ON ANY OTHER BASIS FOR SPECIAL, CONSEQUENTIAL, INCIDENTAL, EXEMPLARY OR INDIRECT DAMAGES IN CONNECTION WITH (OR ARISING FROM) THIS DOCUMENT, WHETHER OR NOT FORESEEABLE AND WHETHER OR NOT PACBIO IS ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Certain notices, terms, conditions and/or use restrictions may pertain to your use of PacBio products and/or third party products. Refer to the applicable PacBio terms and conditions of sale and to the applicable license terms at [pacb.com/license](https://pacb.com/license).

Trademarks:

Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, SBB, Revio, Onso, Apton, Kinnex, and PureTarget are trademarks of PacBio.

PacBio

1305 O'Brien Drive

Menlo Park, CA 94025

[www.pacb.com](https://www.pacb.com)

# Table of Contents

---

<b>Introduction</b> .....	<b>3</b>
<b>Step 1: Evaluate the % Activity plot</b> .....	<b>3</b>
Description of the plot.....	3
Using the Activity plot to evaluate your run.....	3
<b>Step 2: Evaluate the performance of the Sequencing control</b> .....	<b>4</b>
What is the Sequencing control? .....	4
How to prepare the Sequencing control for your run.....	4
How to use the Sequencing control for troubleshooting .....	5
<b>Step 3: Evaluate the productivity metrics</b> .....	<b>7</b>
Productivity .....	7
Using the productivity metrics to evaluate and troubleshoot sequencing performance.....	7
Impact of SMRTbell complex overloading and recommended corrective actions .....	9
Impact of complex underloading and recommended corrective actions.....	10
Causes for high P2%.....	10
<b>Step 4: Evaluate the Read Length Density (RLD) Plot</b> .....	<b>10</b>
Description of the plot.....	10
RLD plot features.....	10
Using the RLD plot to evaluate your run .....	11
<b>Step 5: Evaluate the HiFi Read Length Distribution plot</b> .....	<b>14</b>
Description of the plot.....	14
Using the HiFi Read Length Distribution plot to evaluate your run .....	14
<b>Step 6: Verify that SMRT Link Sample Setup parameters were correctly defined</b> .....	<b>15</b>
<b>Troubleshooting Guidance</b> .....	<b>16</b>
<b>Contacting your service and support team</b> .....	<b>18</b>
<b>Additional information</b> .....	<b>19</b>

# Introduction

---

The [PacBio Revio™ system](#) delivers HiFi long-read sequencing at scale with a high-density 25 million ZMW SMRT® Cell and four independent sequencing stages. Typical sequencing performance for key sequencing applications is detailed in the [Revio system specification sheet](#).

This guide provides information on how to troubleshoot suboptimal performance of Revio system sequencing runs using the Sequencing control and primary run metrics. This guide can help you understand the metrics and how to use them to improve sequencing results.

## Step 1: Evaluate the % Activity plot

---

### Description of the plot

The % **Activity** plot is available **only** during acquisition and provides an early view on sequencing performance **before** final metrics are produced. This plot is found in the Preview Metrics section of the **Run Details** page. The plot displays the percentage of ZMWs that are currently generating a sequencing read of high-quality data (HQ region) at a given time during acquisition; compared to the P1 metric which is the cumulative count of ZMWs that produced a HQ region at **any point** during movie acquisition. The maximum value is typically around 40% and is reached about 5 hours into movie acquisition. A maximum value of 40% is equivalent to loading in the optimal 60-70% P1 range.

### Using the Activity plot to evaluate your run

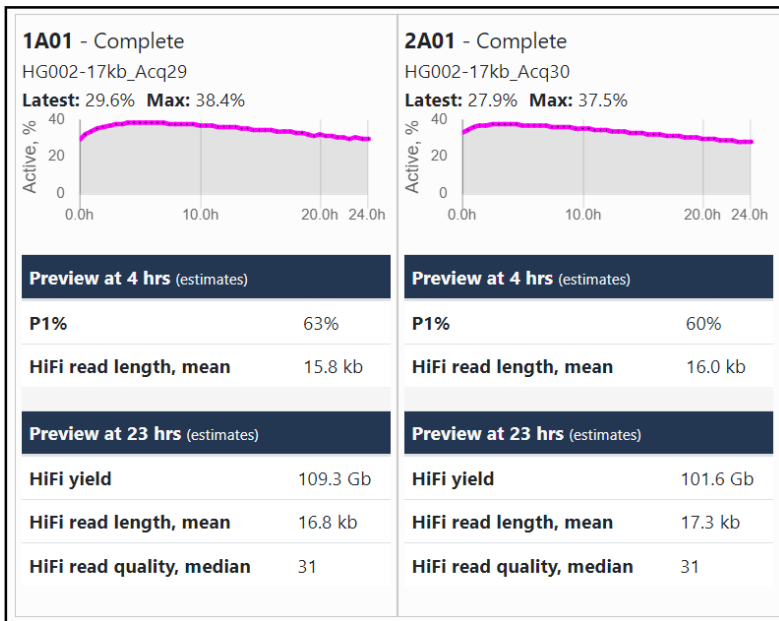
Figure 1 shows the typical pattern expected from a high-quality SMRTbell® library. There are two main items to note for troubleshooting: the rate of decay in active ZMWs and the maximum activity achieved.

The maximum activity achieved can help inform you if adjustments need to be made to the on-plate loading concentration (OPLC) to optimize P1 for your next acquisition of the same sample type.

The rate of decay in active ZMWs is a critical characteristic for identifying problems with sample or consumable quality. A high rate of decay in active ZMWs can be caused by a few different factors. The most prevalent causes are:

- Low sample quality; either excessive DNA damage or a polymerase inhibitor contaminant carrying through both gDNA extraction and bead cleanup to the sample plate.
- Mishandling or incorrect storage of sequencing reagent plates reducing consumable quality, such as storing plates at room temperature, exposure to light, or surface contamination on the foil.

A rare cause of high decay rate on Sequel systems was failure of the oxygen exclusion system. This should be even rarer with Revio systems given the enclosed SMRT Cell; to date we have not observed this issue on the Revio system.



**Figure 1:** Percent Activity plots for two typical human WGS SMRTbell libraries acquired over 24 hours on a Revio system using SMRT® Link v13.0.

## Step 2: Evaluate the performance of the Sequencing control

### What is the Sequencing control?

The Sequencing control is a pre-bound complex that serves as a spike-in control for monitoring sequencing performance. It is added to the polymerase-bound sample at the end of the SMRT® Link Sample Setup procedure. Evaluation of control performance and comparing it to sample performance helps to determine whether to focus troubleshooting on the sample or the sequencing components.

The Sequencing control consists of a stable bound complex containing a defined 11 kb SMRTbell (sequence derived from Lambda), polymerase, an annealed primer, and a nucleotide. Control read filtering occurs on-instrument, where they are separated from sample reads, and output to the `fail_reads.bam` file. Filtering is optimized to minimize false positives, and, as such, a few low-quality control reads **may** pass through the filter to the `HiFi_reads.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 Sequencing control for your run

The Sequencing control is provided as a stock solution and requires a serial dilution prior to adding it directly to the sample. It is **critical** to diligently follow the directions for preparing the control as indicated in SMRT Link Sample Setup. The control should be kept on ice when handling and 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. Make a fresh preparation each time.

## How to use the Sequencing control for troubleshooting

When a sequencing performance issue is observed for a sample, you should first evaluate the sequencing metrics of the Sequencing control within the SMRT Link Runs module. The run metrics table (Figure 2) contains the primary metrics for the control, such as counts and read length, as well as sample.

Well >		Run >		Productivity			HiFi reads				Polym... >		Control reads <				
Plate well	Well name	Status	Movie time	Total bases	P0	P1	P2	Reads	Yield	Length (mean)	Read quality (medi...)	Q30+ bases	Pol. read length (mean)	Reads	Read length (mean)	Concor... (mean)	Concor... (mode)
1 A01	20231023-2_8...	Complete	24 hr	1,334 Gb	40%	60%	0%	7.5 M	128.5 Gb	17.1 kb	Q33	94%	89.1 kb	3,183	65.1 kb	91%	91%
1 B01	20231023-2_8...	Complete	24 hr	1,244 Gb	46%	53%	0%	7.1 M	119.5 Gb	16.9 kb	Q33	94%	92.6 kb	3,014	70.9 kb	91%	91%
1 C01	20231023-2_8...	Complete	24 hr	1,009 Gb	43%	57%	0%	5.3 M	100.4 Gb	18.9 kb	Q33	92%	70.7 kb	2,640	60.1 kb	91%	93%
1 D01	20231023-2_8...	Complete	24 hr	1,108 Gb	42%	58%	0%	5.8 M	110.3 Gb	18.9 kb	Q34	93%	75.9 kb	2,684	66.0 kb	91%	93%

**Figure 2:** Sequencing metrics for a 4 SMRT Cell Revio run as shown in the SMRT Link Run Details page.

Run Design parameters: Two Kinnex full-length RNA and two Kinnex 16S rRNA SMRTbell libraries; Revio SMRT Cell tray (P/N 102-202-200); Revio polymerase kit (P/N 102-817-600); Revio sequencing plate (P/N 102-587-400); SMRT Link v13.0; with adaptive loading; 24 hours acquisition.

The expected value for each metric, at various supported movie lengths, are shown in Table 1. The expected values are averages from all sites and reflect typical expectations from a correctly-prepared control and real-world sample.

- **Reads:** The total number of control reads detected by the primary analysis software.
- **Read length (mean):** The average polymerase read length of the control reads, including insert, adapters, and all passes around the SMRTbell template.
- **Concordance Mean/Mode:** The percent agreement of the control raw reads to the control reference sequence. (This does not use consensus base calling.)

Movie Length (hrs)	Reads		Read length (kbs)		Concordance	
	Median	Quartile Range, 25%-75%	Median	Quartile Range, 25%-75%	Median	Quartile Range, 25%-75%
<b>12</b>	1,710	900 – 2,400	33.4	30.4 – 36.5	0.909	0.906 – 0.912
<b>24</b>	1,730	1,100 – 2,400	59.6	53.7 – 65.2	0.907	0.904 – 0.909
<b>30</b>	1,550	1,200 – 2,100	64.6	59.3 – 69.7	0.905	0.902 – 0.906

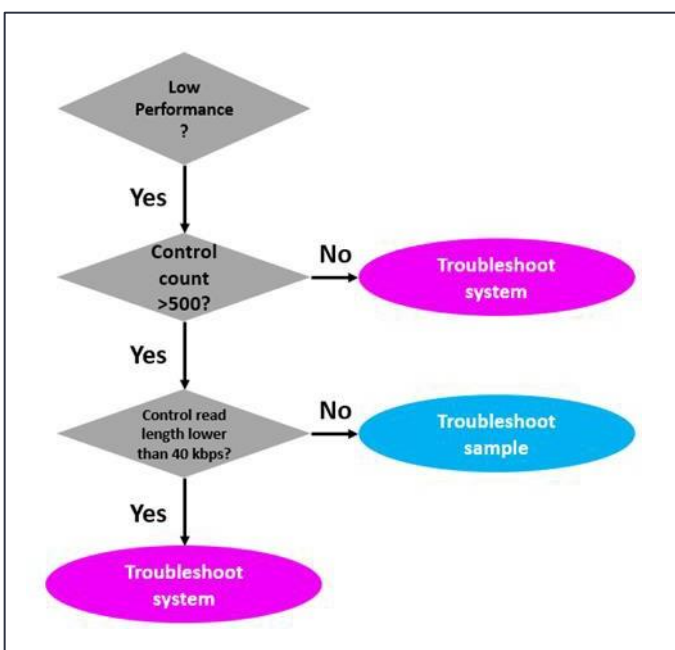
**Table 1:** Average Control reads, read lengths and concordance for common movie lengths on Revio.

Run Design parameters: Revio SMRT Cell tray (P/N 102-202-200); Revio polymerase kit (P/N 102-817-600); Revio sequencing plate (P/N 102-587-400); HiFi Reads application; SMRT Link v13.0.

To successfully interpret these metrics, the acquisition run parameters (such as movie length, OPLC, and so on) must also be considered. Some important considerations for the Control read count and Control read length are:

- The Control read count is typically 1,000-3,000 and is independent of movie time.
  - The Control read length is **strongly dependent** on movie time (see Table 1). For high-quality libraries, the Polymerase read length for the sample is similar to and often higher than the control read length.
  - The Control counts **negatively** correlate with increasing OPLC of the sample due to competition between molecules of sample and control for binding sites within the ZMWs of a SMRT Cell.
  - Too high an OPLC can result in overloading and decreased Control Read Length 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 4 and 5).

- Good control performance indicates proper functioning of the Revio system and consumables. Troubleshooting should focus on the upstream workflow: gDNA extraction, library preparation and/or ABC (annealing, binding and cleanup).
- Low Control read count (<500) may indicate poor loading due to consumables mishandling or instrument error. If a run has low P1% and low control read count, contact PacBio Service and Support to investigate. While poor loading for sample and control may indicate a consumable or instrument issue, it may **also** be due to improper library preparation or a contaminant in the sample that inhibits the immobilization of the control.
- High control read count (>20,000) typically indicates an issue with dilution or handling in Sample Setup. This can be due to miscalibrated pipettors, non-homogenous solutions, or deviations from protocol.
- If a run had good Control read length (>40 kb for 24 hour movies) but poor Polymerase read length, then troubleshooting should be focused on library quality, such as extraction and/or library preparation.
- Low Control read length (<40 kb for 24 hour movies) is typically linked with low sample Polymerase length.
  - In some cases, this is due to overloading (P1>75%). In such cases, we recommend using adaptive loading to prevent overloading or to lower the OPLC.
  - In other cases, both control and sample read length are impacted by sample contaminants. If you consistently see poor read lengths, troubleshooting should be focused on sample and library preparation.
  - Finally, low Control read length may indicate issues with consumable handling or storage, or instrument error. If you see occasional low read length, contact PacBio Service and Support to investigate the affected runs.
- If the Control Reads and the Control Read Length were both **lower** than expected, then troubleshooting should consider the following: storage of the sequencing consumables; carry over of cleanup beads into the sample plate (**no** amount of cleanup beads should ever be present in the sample wells of the sequencing reagent plate.) Contact PacBio Service and Support to look at the instrument.
- If both the Control Reads and Control Read length 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 3:** Flow chart for Sequencing control metric troubleshooting.

## Step 3: Evaluate the productivity metrics

---

### Productivity

The SMRT Link Run Details table (see Figure 2) includes primary sequencing metrics for productivity, commonly referred to as **loading metrics**. The productivity metrics show how many ZMWs on the SMRT Cell have produced useful sequencing data and if the correct amount of sample was used. After completion of sequencing data acquisition for a given SMRT Cell, the Revio system's primary analysis pipeline classifies each ZMW (as well as 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 present but no HQ region was detected.

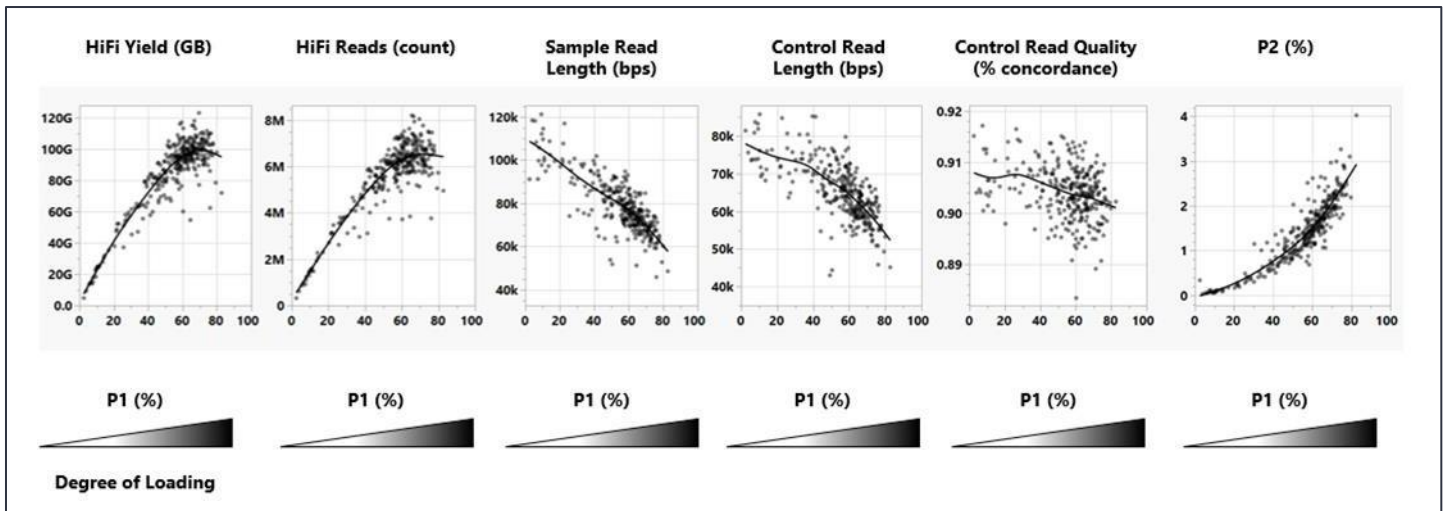
These metrics are represented as a percentage of the total number of ZMWs present on a SMRT Cell.

- The **P0** metric indicates the number of ZMWs that had **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  $\geq 50$  bases of high-quality sequencing (i.e. a HQ Region) at **any** time during acquisition will be classified as **P1**. The reads from P1 ZMWs that generate a HiFi read (consensus accuracy  $\geq QV20$ ) are stored in the `HiFi_reads.bam` file and are used in downstream applications.
- All other ZMWs producing detectable signal, but not a HQ region, are classified as **P2** and **not** output by the system (this includes P0s). These ZMWs commonly have high and noisy background signal and may contain multiple active sequencing complexes per ZMW. It is a common misconception that P2 refers to doubly loaded ZMWs. This is not strictly correct; however, most P2 ZMWs are of this variety.

### Using the productivity metrics to evaluate and troubleshoot sequencing performance

When evaluating the performance of a sequencing run, assess the loading of the sample SMRTbell complex by primarily considering the P1 metric. The P1 metric will generally provide an 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 analyses. As shown in Figure 4, once an optimal inflection point (60-70% P1) of loading is reached, a higher degree of loading results in poorer raw sequencing performance, as read lengths, yield and concordance metrics decline. Note that P1 and P2 still increase past this point, however, on average, the reads are slightly shorter, of lower raw accuracy, with HiFi conversion suffering. Also, note that the performance of the Sequencing control itself is negatively affected.





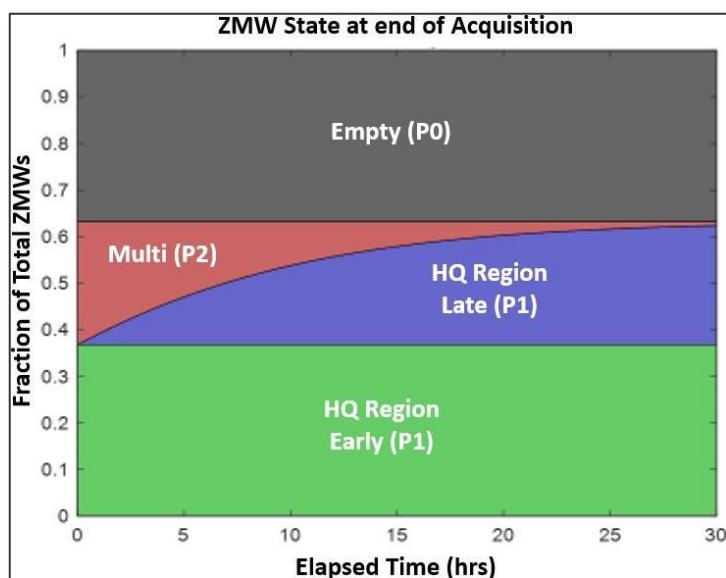
**Figure 4:** Representative sequencing performance metrics, as a function of loading, from a customer who runs human WGS projects on their Revio system. Run Design criteria: Revio 25M SMRT Cells (P/N 102-202-200); Revio polymerase kit (P/N 102-817-600); Revio sequencing plate (P/N 102-587-400); HiFi Reads application; ICS  $\geq$  v12.x; insert size 13,000 to 23,000 bps; 24 hours acquisition; SMRTbell prep kit 3.0 (P/N 102-182-700).

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 (P1) state increase (see Figure 4). The time taken for a P2 state ZMW to transition to a P1 state ZMW (a single remaining polymerase still sequencing) also lengthens. This results in far fewer P1 state ZMWs using the full length of the acquisition time for useful, high quality, long polymerase read length sequencing.

Conversely, a low P1 (<50%) results in lower HiFi Yield, as fewer actively sequencing complexes are present on the SMRT Cell. Mean polymerase read length may be slightly longer as fewer multi-loaded ZMWs (Multi P2 state Figure 5) are present from the beginning of acquisition under such conditions.

Determining the optimal loading range depends on balancing these interactions with the overall goal of the sequencing project. For most applications, such as Kinnex or *de novo* genome assembly, where many HiFi reads of long insert libraries are the priority, a SMRT Cell run with a P1 in the range of 50-70% is optimal.





**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 read with a HQ region. **Green** indicates ZMWs that produced useful, high quality sequencing, immediately upon acquisition start (“HQ Region Early (P1)”) and tend to produce the longest read lengths with the most HiFi bases. **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 and fewer HiFi bases.

## Impact of SMRTbell complex overloading and recommended corrective actions

Overloading polymerase-bound complexes on a SMRT Cell will result in shorter Polymerase Read Length of the sample and thus decrease the HiFi yield (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 polymerase read length of the sample. For most types of long-read sequencing, a slightly **smaller** number of **longer** polymerase reads results in **higher** performance in secondary analysis, compared to a larger number of shorter polymerase reads.

**How to tell if a SMRT Cell is overloaded?** With increasing OPLC, P1 and P2 will increase, while P0 will decrease, though not necessarily linearly. Once P1 increases to >80%, P2 increases beyond 5%, and the mean Polymerase Read Length of the sample has decreased significantly, the SMRT Cell has most likely entered an overloaded state. P2% is a good secondary metric for assessing overloading, however, it is not as sensitive as P1 or polymerase read length. Sample contamination or SMRT Cell surface defects could inflate P2 levels in the absence of overloading (see below). A high P2, taken by itself, should **not** be automatically interpreted as overloading, but may indicate other underlying issues. If an individual SMRT Cell has shorter-than-expected Polymerase Read Length with lower-than-desired HiFi Yield, a P1 value >80%, a P2 value >5%, and sample quality is not in doubt, then it is generally recommended that you rerun the sample at a lower OPLC to increase performance. However, 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 underloading and recommended corrective actions

One of the most common, and easiest to fix, causes of low P1 reads and HiFi Yield is underloading of polymerase-bound complexes onto the SMRT Cell. If the P1 for a SMRT Cell is substantially lower than 40%, P0 and P2 values are both low, and sample and control read lengths and control counts are as expected, then the SMRT Cell was likely underloaded.

To correct this, and increase HiFi yield, try rerunning the sample at a higher OPLC (but still within our recommended range). As complex loading on SMRT Cells is not a linear process, relatively large increases in OPLC (1.2-2x) are usually required to see an impact on the productivity loading profile.

Another common cause of underloading includes inefficient generation of polymerase-bound complexes due to operational mistakes during annealing, binding and cleanup (ABC). A useful indicator that can sometimes point to bound complex-related quality issues is shorter than expected polymerase read length on an underloaded SMRT Cell with normal control performance. If there is any doubt regarding the generation of the bound complex, then first consider redoing ABC **before** attempting to rerun another SMRT Cell.

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 and base rate for the sample. These are typically severely reduced. Such samples usually showed gDNA degradation on molecular weight sizing gels and/or had unexpected spectrophotometric absorbance ratios prior to library preparation. Rerunning a “bad” sample at a higher OPLC than what we recommend will **not** rescue that sample. The quality of the sample should be evaluated, and likely reextracted with additional purification steps and/or care to limit degradation and DNA damage.

### Causes for high P2%

A high P2 value can have several causes, although the most common is from overloading the SMRT Cell which can be easily remedied by lowering the OPLC or increasing the movie length (see Figure 5). Historically, high P2s were commonly caused by excess polymerase mistakenly added, or carried over, during complex binding (e.g. a dilution step was missed) or poor cleanup performance. Our current bead cleanup technology is very effective at removing excess polymerase and this is rarely seen in practice these days.

## Step 4: Evaluate the Read Length Density (RLD) Plot

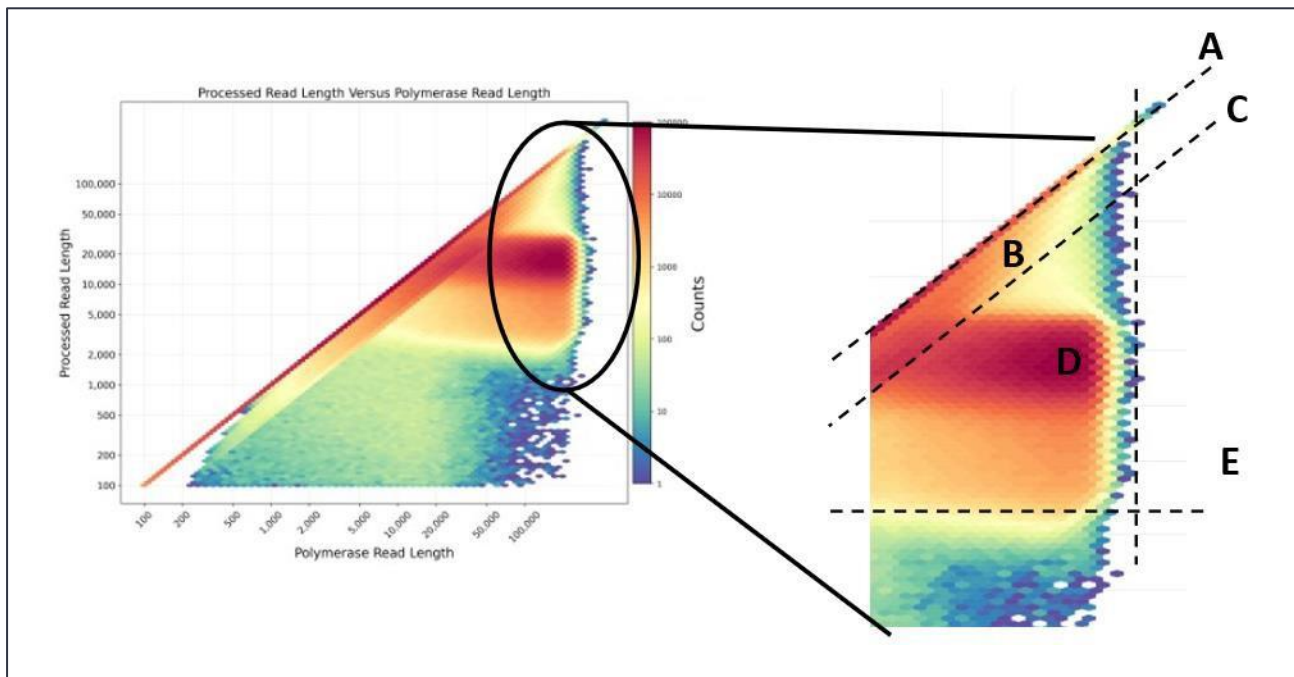
### Description of the plot

The RLD plot, found on the SMRT Link Run Details page, is a useful tool for visualizing and evaluating the distribution of SMRTbell insert sizes and polymerase read lengths in your sequencing dataset (see Figure 7). Each dot represents the amount of P1 ZMWs according to the polymerase read length determined by the HQ region finder on the x-axis, and an accurate estimate of the insert read length on the y-axis (both axes are logarithmic.) The density of reads is color-coded, with **purple** representing low density of P1 counts, and **red** indicating high density.

### RLD plot features

- **Polymerase Read Length:** The complete length of the read, including all adapter, barcode and insert bases across both strands from all passes around the template. **Note:** The axis is logarithmic.
- **Processed Read Length:** The longest insert length detected between two adapter sequences per ZMW. **Note:** The axis is logarithmic.
- **Line A:** The primary diagonal; represents early polymerase terminations along the template before or at the first observed adapter (e.g. from DNA damage).

- **Region B:** The area between lines A and C; reads that terminated after the first observed adapter but before the second observed adapter.
- **Line C:** The secondary diagonal; terminations at or near the second observed adapter.
- **Region D:** The primary HiFi generating region representing reads >2 passes across the template.
- **Line E:** The size selection boundary cut-off (if applicable).



**Figure 6:** RLD plot taken from a Revio system acquisition of a bead size-selected human WGS library, with MW mode length between 15-18 kbps (MW trace not shown), overlaid with useful troubleshooting features (see text for description). As the Polymerase Read Length is always  $\geq$  Processed Read Length, for a given ZMW, no data appears to the left of the primary diagonal.

## Using the RLD plot to evaluate your run

RLD plots are extremely useful for interpreting basic characteristics of your SMRTbell library to aid in troubleshooting. The plot 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 (DNA damage); proportion of reads entering CCS and generating HiFi data; and other information.

The primary diagonal line (**Line A, Figure 6**), running in the center of the plot, corresponds to reads where the Polymerase Read Length = Processed Read Length. Reads which fall along this line are reads in which **no** adapter was detected in the acquired HQ region. Data from these reads are exclusively generated from a single pass of the polymerase across the insert and will **not** generate HiFi data. A high density of reads appearing along this diagonal usually indicates a sample quality or library preparation issue. This is caused by the inability of the polymerase to reach an adapter in the first observed pass across the insert. If enrichment is located about the expected template size range or shorter, then typical root causes include: DNA damage (including nicks); poor ligation of adapters (polymerase falls off the end of the template); contamination by polymerase inhibitors; or failure to use Pre-extension. If DNA damage or contaminants are suspected, then additional AMPure PB bead purifications of the starting gDNA or re-extraction using a different method are good strategies for successful sequencing.

If enrichment is located **longer** than the expected template size range, then the ligation reaction should be suspected first. Missing molecular adapters due to chimeras (which are end-to-end ligations of DNA inserts with

adapters on both ends) or incorrect adapter-to-template ratios could be to blame due to having too little or unannealed adapters, or too much template during ligation. Make sure to quantify the template **correctly** so that an accurate amount of template is added to the reaction. On rare occasions, a “sticky” SMRT Cell with a high background signal, could cause a buildup of reads along this diagonal. These reads are much longer than the expected insert length, are highly repetitive and can be confirmed as “junk reads” by their inability to map to a suitable reference.

When the Polymerase Read Length becomes longer than the Processed Read Length, reads accumulate to the right of the primary diagonal and spread out along the x-axis. These are reads that are entering their second or higher observed pass along their insert, i.e. have detected  $\geq 1$  adapter. There are many important features in this region.

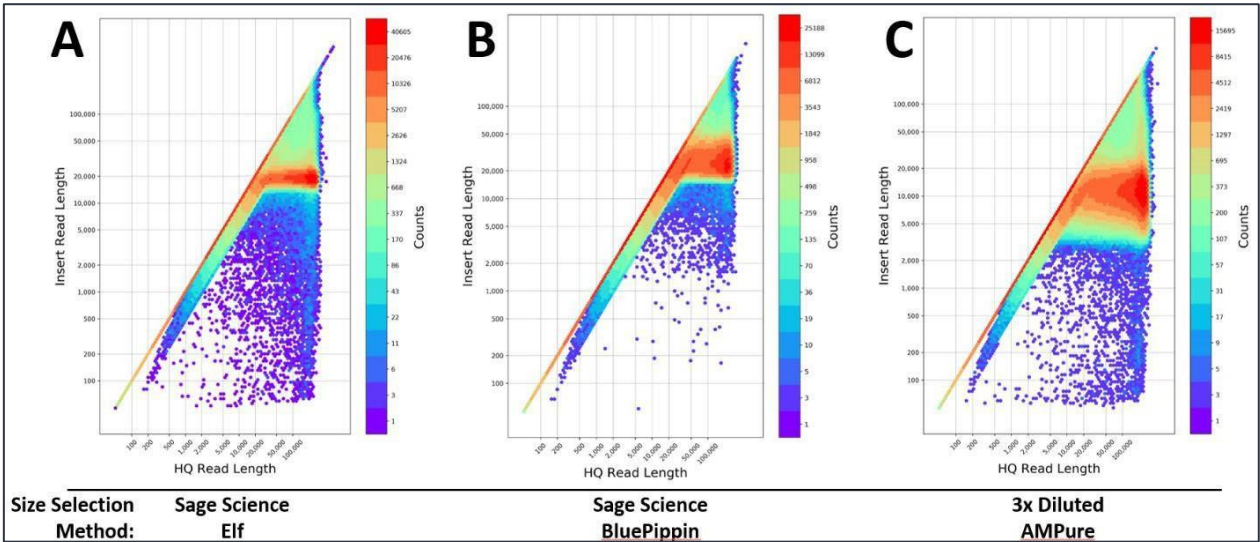
A secondary diagonal line (**Line C, Figure 6**), shifted to the right and parallel to the primary diagonal, may also form and represents polymerase read lengths of 2x the SMRTbell insert size (as determined by adapter detection per molecule). This secondary diagonal represents termination of polymerases at or near the second observed adapter. This line usually represents a small proportion of the total reads in a dataset. It is rare to see an accumulation of reads on this line.

The data that falls between lines A and C (denoted as **Region B, Figure 6**), are those polymerases that sequenced past one observed adapter but terminated sequencing before the second observed adapter. Accumulation of reads here is typically caused by DNA damage.

If a polymerase has sequenced  $>2$  observed adapters, the data point will fall to the right of Line C. This region indicates circular consensus sequencing (CCS), and is where HiFi reads are generated (**Region D, Figure 6**). It is normal to find a high density of reads in this region, at polymerase read lengths multiple times the expected insert size, and with a processed read length distribution (y-axis) matching that of the SMRTbell library. All sample types, including short amplicons, Iso-Seq and HiFi libraries, are expected to enter CCS mode and show a large accumulation of reads in region D. In CCS mode, the polymerase is very stable and will produce the longest HQ regions. Reads in CCS terminate mostly because of laser-induced photodamage, or because movie acquisition ended.

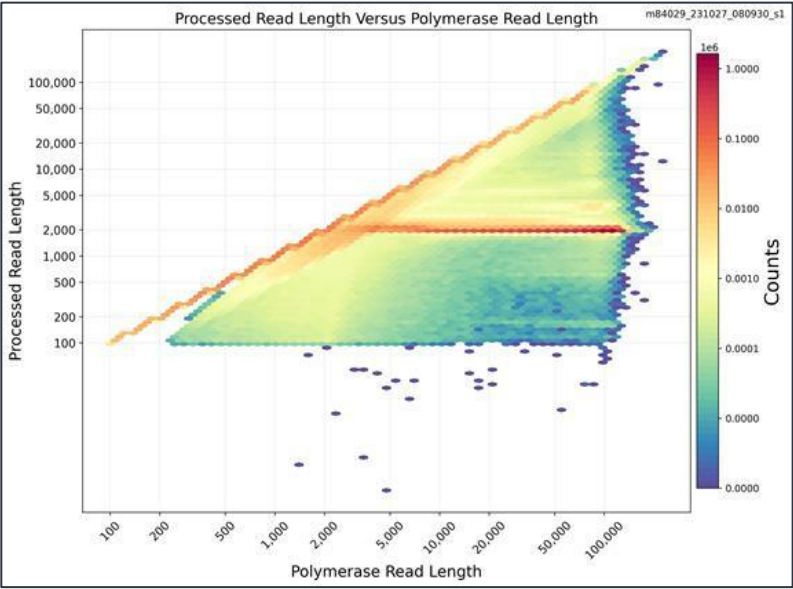
A third prominent line, found in size-selected libraries, corresponds to the size selection cutoff or depletion boundary (**Line E, Figure 6**) below which molecules shorter than this length were removed from the SMRTbell template population. It is denoted by a sharp drop in read density below a line parallel to the x-axis at the Processed Read Length of the size selection cutoff. If a strict size selection cutoff (e.g. Sage Science BluePippin size selection) was used during library preparation, then this boundary should be very well defined. Whereas, a bead-based or SRE size selection method, which works to gradually deplete molecules over a wide size range, will have a more diffuse boundary. If this line is **not** well defined, is at the wrong length, or the HiFi read length is **not** of the expected size, then the size-selection step should be suspect.

It is important to remember that the pattern and interpretation of this plot is very much dependent on the application, how the template was prepared, and the Run Design parameters used for data collection (such as pre-extension and movie length). For example, a gDNA library used for de novo assembly with a 30 hour movie (Figure 7) will have a very different density distribution profile compared to a 2000 bp PCR amplicon library acquired over only 12 hours (Figure 8). Additionally, the various size selection methods available will greatly influence this plot (see Figure 6). Interpretation of the RLD plots is a valuable tool for troubleshooting sequencing runs and SMRTbell libraries.



**Figure 7:** RLD plots of HiFi SMRTbell libraries size-selected by different methods and run on Sequel II 8M SMRT Cells. **A.** Human library size selected using a Sage Elf system showing an ideal, relatively narrow, insert size range (note distribution along the y-axis). **B.** Plant library size-selected using a Sage BluePippin system showing a discrete size cutoff on the low end, but a broader range of longer fragments. **C.** Human library size-selected using the 3x Diluted AMPure method showing the broadest and shortest insert size range.

Run Design parameters: Sequel II SMRT Cell tray (P/N 101-389-001); Sequel II binding kit 2.0 (P/N 101-789-500); Sequel II sequencing 2.0 kit (P/N 101-820-200); SMRT Link v8.0; 30 hours acquisition.



**Figure 8:** RLD plot of a 2000 bp SMRTbell library derived from a PCR amplicon. A very discrete insert size was observed. Run Design parameters: Revio SMRT Cell tray (P/N 102-202-200); Revio polymerase kit (P/N 102-817-600); Revio sequencing plate (P/N 102-587-400); SMRT Link v13.0; with adaptive loading; 12 hours acquisition.



## Step 5: Evaluate the HiFi Read Length Distribution plot

### Description of the plot

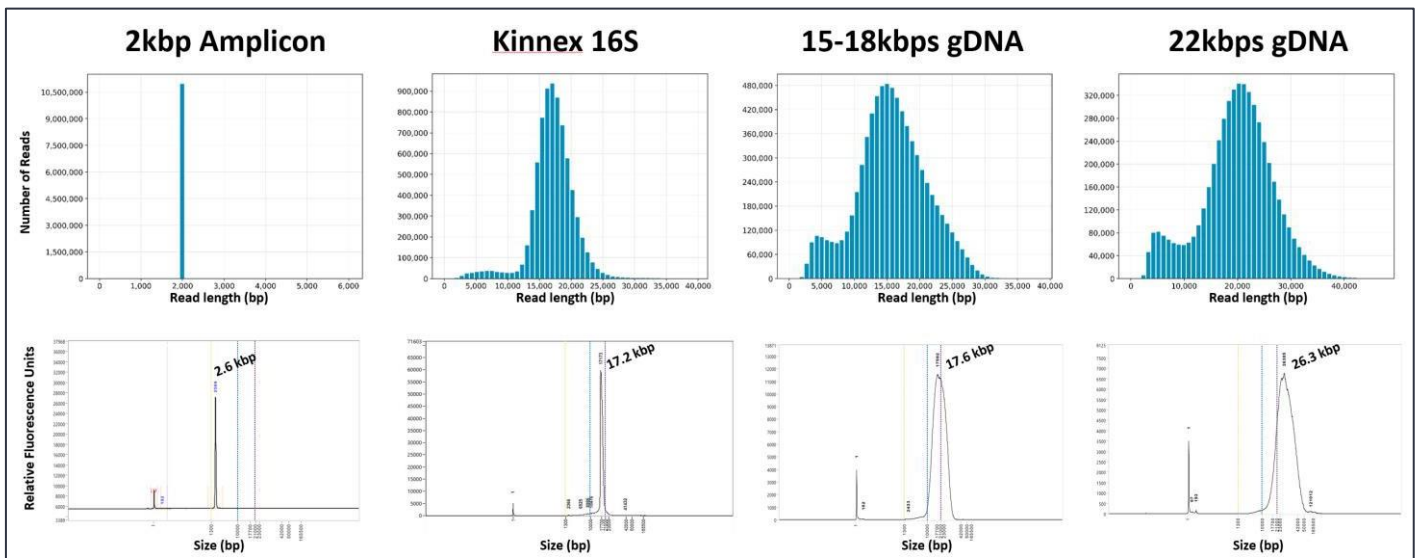
The **Read Length Distribution plot** is available for each acquisition in the Plots section of the SMRT Link Run Details page. It shows the distribution of the lengths of each consensus HiFi read produced per ZMW. It is a good estimate of the fragment size distribution of the original SMRTbell library as the length of each HiFi read is determined by multiple passes around the template and the detection of multiple adapter sequences.

**Note:** The HiFi read length does **not** include adapter sequence, but **does** include barcode/index sequence.

### Using the HiFi Read Length Distribution plot to evaluate your run

The mode and distribution of the HiFi read lengths are very much library and application specific and should be evaluated for your specific sample type. In Figure 9 we see various SMRTbell library types, their MW sizing traces by Femto Pulse, and the resulting HiFi read length distributions. The read length mode typically reflects the mode observed in pulse field gel-based MW sizing traces, but may be slightly shorter. This is normal and not of concern to data quality or final application performance.

Additionally, there may be a slight enrichment of HiFi reads around 5 kb. This is due to the slightly favored immobilization of shorter molecules into ZMWs during diffusion, but is only observed on non-size-selected libraries with high fractions of molecules under 7 kb. Even extensive amounts of these shorter HiFi reads do not impact downstream application performance (data not shown). However, it will reduce overall HiFi GB yield as this metric is a product of the number of HiFi reads times the HiFi read length. Shorter HiFi read lengths will, therefore, lead to lower expected HiFi GB yield. These shorter molecules can easily be removed by gel-based size selection (such as BluePippin) or greatly depleted by a combination of SRE and bead-based size selection. Any major shift in mode or distribution observed is usually related to poor performance, or lack of size selection method, and troubleshooting should be directed there first.



**Figure 9:** Typical HiFi Read Length Density plots of 4 SMRTbell libraries from different applications acquired over 24 hours on a Revio system with SMRT Link v13.0. Also shown are the Femto MW traces representing those same libraries acquired on Agilent Femto Pulse instrument.

## Step 6: Verify that SMRT Link Sample Setup parameters were 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 seen for a sample, verify that the Sample Setup sheet was **correctly** generated for the specific sample in question. Common mistakes can be identified and fixed **before** the same bound sample complex is to be run again.

- Pay close attention to any warning messages that may appear in Sample Setup. For example, if Sample Setup warns that the sample concentration is out of range, seriously consider concentrating or diluting the sample with 1x SMRTbell cleanup beads or, prepare additional SMRTbell libraries. Ignoring these warnings is a very common cause of low P1.
- Select the **correct** Library type for your SMRTbell library. Selecting the incorrect type will prevent accurate adapter-calling and will lower system performance. Specifically, you will see a very large increase in the **Missing Adapter** metric in the Run Details page.
- The polymerase should be diluted fresh for each run, kept on ice while in use, and diluted polymerase should **not** be stored and reused.
- 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 gDNA 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 >10 kb, PacBio recommends using a pulsed-field-based electrophoresis method. For assessing shorter libraries, the Agilent Bioanalyzer system is a good choice. We do **not** recommend the TapeStation as it tends to be inaccurate and routinely overestimates fragment sizes.
- The **Concentration** and **Insert Size** fields are used to automatically calculate the molarity of the SMRTbell library. Accurate molarity determination is important for generating optimized reaction conditions for primer annealing, polymerase binding and sample immobilization on the SMRT Cell.
- Ensure that the latest version of SMRT Link is installed. 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.
- If requesting assistance from PacBio Service and Support, the Sample Setup sheet can easily be exported by using your browser's **Print** and **Save as PDF** function.



# Troubleshooting Guidance

<b>Sample shows poor sequencing performance (Sequencing control performance is normal)</b>		
<b>Symptom</b>	<b>Potential causes</b>	<b>Possible actions/solutions</b>
<b>Sample shows short Polymerase Read Length</b>	Excessively high sample OPLC (on-plate loading concentration).	Reduce the sample OPLC or turn on adaptive loading.
	Sample quality issue - such as highly fragmented DNA, high amount of DNA damage, presence of a contaminant.	Confirm sample QC; repurify the sample. Consider changing methodology.
	Inefficient adapter ligation reaction.	Verify adapter ligation conditions used during SMRTbell library construction; reperform library preparation step if needed.
<b>Sample shows low P1 productivity metric</b>	OPLC was too low.	Increase OPLC to within our recommended range.
	Ignored sample concentration warning in Sample Setup.	Dilute library with elution buffer.
	Manual pipetting error during primer annealing, polymerase binding or complex cleanup (ABC) steps.	Redo sample ABC steps.
	Sample quality issue, such as highly fragmented DNA, high amount of DNA damage, presence of a contaminant.	Confirm sample QC (DNA molecular weight, purity and concentration); repurify the sample. Consider changing extraction methodology.
<b>High P2 productivity</b>	Excessively high sample OPLC.	Reduce OPLC. Extend movie time.
	Sample quality issue, such as fluorescent contaminant.	Confirm sample QC (DNA purity); repurify the sample. Consider changing extraction methodology.
	Excessively high unbound polymerase carryover.	Redo ABC with particular care at bead cleanup 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.
<b>HiFi Read Length Distribution plot shows longer than expected library insert size</b>	Concatemers present due to inefficient adapter ligation reaction.	Verify adapter ligation conditions used during SMRTbell library construction; redo the library preparation step if needed.
	Sample MW issue, such as incorrect library insert size.	Confirm sample QC and molarity (DNA molecular weight); redo the sample size selection step.
<b>HiFi Read Length Distribution plot shows increased "5kb shoulder"</b>	High amount of <7kb fragments present in gDNA pre or post shearing.	Perform SRE on gDNA pre-shearing; Perform gel-based size selection on library; Revisit sample storage and extraction.

<b>Sequencing control shows poor sequencing performance</b>		
<b>Symptom</b>	<b>Potential Causes</b>	<b>Possible Actions/Solutions</b>
<b>Low Control reads (Sample sequencing performance is either normal or poor)</b>	Pipetting error during preparation of Sequencing control.	Reprepare the Sequencing control using calibrated pipettes.
	Sample quality issue, such as contaminant.	Confirm sample QC (DNA purity). Repurify the sample. Consider changing extraction methodology.
	Excessively high sample on-plate loading (OPLC) concentration.	Reduce sample OPLC.
	SMRT Cell or instrument hardware issue.	Contact PacBio Service and Support.
<b>Short Control read length (Sample sequencing performance is poor)</b>	Magnetic bead carryover to sample plate.	Ensure that there is zero bead carryover during cleanup.
	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, such as contaminant.	Confirm sample QC (DNA purity), Repurify the sample. Consider changing extraction methodology.
	SMRT Cell or instrument hardware issue.	Contact PacBio Service and Support.
<b>High Control read count (Sample sequencing performance is normal or poor)</b>	Pipetting error during preparation of Sequencing control.	Reprepare the Sequencing control using calibrated pipettes.
	A dilution step was inadvertently omitted during preparation of the Sequencing control.	Reprepare the Sequencing control.
	Excessively low sample OPLC.	Confirm molarity. Increase sample OPLC.
	Instrument hardware issue.	Contact PacBio Service and Support.

# Contacting your service and support team

---

If you are unable to improve sequencing performance or suspect an issue with PacBio consumables or instruments, please reach out to your local Field Application Scientist (FAS) or Field Service Engineer (FSE) for assistance by filing a case via email or calling the support line (which is answered by Global Service and Support in Menlo Park, California) here:

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

To speed the support process, 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. (See your version of the SMRT Link user guide [here](#).)
  - If not available: Run ID, movie ID, date, sample well, and instrument serial number (84XXXX).
  - Include any PDF Reports of previous runs for the same sample, SMRTbell library or bound complex.
- Sample information:
  - What organism(s)/tissue(s) were the samples from?
  - Have you successfully sequenced from similar samples before?
  - What extraction protocol was used?
  - What PacBio library preparation Procedure & Checklist did you follow? (P/N and version)
  - Attach any QC of the extraction preparation or library preparation that you performed.
    - Was Qubit quantification performed on the starting material and did it correspond to the Nanodrop quantification?
    - Include any MW sizing gel images/traces, including RIN numbers if relevant.

## Additional information

---

- Resources for best practices during sample preparation, SMRTbell preparation protocols, and software documentation can all be found on [www.pacb.com](http://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.
- Sample and library preparation [protocols](#) for all applications.
- Helpful information on sample extraction best practices can be found [here](#) and [here](#).
- Useful [table](#) of experimental design criteria, protocol links, input requirements and SMRT Link workflows.
- A similar [table](#), but for Nanobind extraction kits.
- [Product shelf-life and storage recommendations](#).
- SMRT Link software [user guides and release notes](#).
- PacBio glossary of terms: See the Help section of SMRT Link.
- The sequence of our Sequencing control can be requested from [support@pacb.com](mailto:support@pacb.com).

Research use only. Not for use in diagnostic procedures. © 2024 Pacific Biosciences of California, Inc. ("PacBio"). All rights reserved. Information in this document is subject to change without notice. PacBio 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 PacBio products and/or third-party products. Refer to the applicable PacBio terms and conditions of sale and to the applicable license terms at [pacb.com/license](http://pacb.com/license).

Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, SBB, Revio, Onso, Apton, Kinnex and PureTarget are trademarks of PacBio.