Sample Quality – Effects of Contaminants on SMRTbell™ Library Preparation and Sequencing

Kristi Spittle, Susana Wang and Primo Baybayan
Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025

Introduction

Sample quality is crucial to successful SMRTbell™ library preparation and must be considered before sequencing with the PacBio® *RS.* One important factor in quality sequencing is accurate concentration measurements. Over-estimating concentration can lead to excess polymerase in the binding reaction and result in poor sequencing. There are many contaminants that can be present in biological samples. Here we will present different contaminants and their effects on quantitation, library preparation, and sequencing.

Methods

A fixed-size, contaminant-free, 11 kb annealed SMRTbell™ library was prepared using the PacBio® *Procedure & Checklist – 10 kb Template Preparation and Sequencing*. An initial 0.45x AMPure® purification was done (per the procedure) to remove excess primer. Aliquots of the sample were prepared and then spiked with varying contaminants (*Figure 1*) at 1%, 10%, and 50% volumes. A separate tube of 100% contaminant was also prepared.

Concentration of each contaminated sample was measured using both NanoDrop® 1000 spectrophotometer (Thermo Fisher Scientific, Inc.) and Qubit® dsDNA HS Assay Kit and the Qubit® 2.0 Fluorometer (Invitrogen). Measurements were taken following manufacturer protocols[1,2] and done in triplicate. Concentrations were determined using undiluted sample for the NanoDrop spectrophotometer and 1:10 sample dilution for the Qubit fluorometer. After quantitation, samples were annealed with primer and bound with polymerase using the standard PacBio® procedures[3]. Samples were initially sequenced using the Standard Seq v2 sequencing protocol. Underperforming samples were run again using both Standard Seq v2 and MagBead Seq v1 sequencing protocols.

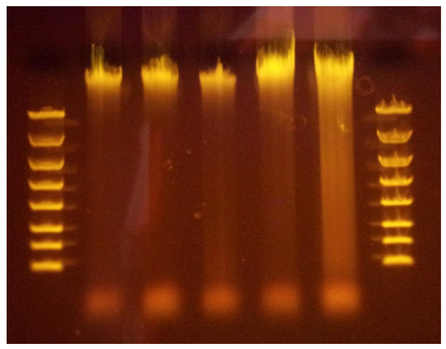
A second set of samples, from various in-house projects, were analyzed for sample quality prior to SMRTbell™ library construction. Concentration measurements and gel electrophoresis results (FlashGel® DNA cassette 1.2% - Lonza Group Ltd.) were assessed to determine sample quality.

Results

Sample ID	Qubit	NanoDrop			[ND]-[Qubit]	Fold diff b/w
	[ng/uL]	[ng/uL]	260/280	260/230	Reading	[ND] and [Qubit]
Sample A	89.4	44.7	1.9	4.83	-44.7	2.0
Sample B Strain 1	476	841.5	1.84	1.62	365.5	1.8
Sample B Strain 2	1160	985.2	1.85	1.68	-174.8	1.2
Sample B Strain 3	558	1283.9	1.78	1.13	725.9	2.3
Sample B Strain 4	214	1064.7	1.76	1.13	850.7	5.0
Sample B Strain 5	624	1118.3	1.86	1.81	494.3	1.8
Sample C	60.8	301.5	1.89	1.24	240.7	5.0
Sample D	69.4	728.7	1.81	1.36	659.3	10.5
Sample E	40.6	387.9	1.86	1.53	347.3	9.6
Sample F	684	727.8	1.54	0.74	43.8	1.1
Sample G	574	1555.4	1.97	2.31	981.4	2.7

Figure 1. Concentration measurements from 11 in-house project samples.

Results



1.2% Lonza FlashGel® cassette
1 = Lonza Flash Gel Ladder (0.1 – 4K)
2 = Sample B Strain 1
3 = Sample B Strain 2
4 = Sample B Strain 3
5 = Sample B Strain 4
6 = Sample B Strain 5
7= Lonza Flash Gel Ladder (0.1 – 4K)

Figure 2. Gel image of the five strains of Sample B containing contaminants.

		NanoDrop			Qubit
Contaminant	% Contaminant	Concentration	A260/280	A260/230	Concentration (ng/ul)
None	0%	74.01	2.03	2.20	99.1
ssDNA	1%	84.00	1.85	1.57	98.67
	10%	70.54	2.05	2.17	98.27
RNA	1%	96.24	2.06	2.20	100.67
	10%	193.50	2.08	2.31	99.33
Phenol/ Chloroform	1%	5077.08	1.26	1.14	104.67
	10%	5045.20	1.10	1.12	103.33
Glycogen	1%	57.38	2.30	2.76	111.00
	10%	52.83	2.37	2.45	82.00
Ethanol	1%	74.52	2.03	2.18	84.20
	10%	76.38	2.00	2.19	84.07
DMSO	1%	75.33	1.97	0.22	81.00
	10%	72.69	2.00	0.15	82.87
Betaine	1%	75.92	2.02	1.03	97.47
	10%	65.18	2.08	0.36	89.07
SDS	1%	92.15	2.02	2.20	111.00
	10%	71.84	2.01	1.51	122.33
Tween	1%	77.13	1.94	0.87	110.67
	10%	75.22	1.87	0.46	109.33
Triton X 100	1%	1500.14	0.43	0.36	105.33
	10%	4186.96	0.88	0.94	107.33
Heparin	1%	75.58	2.11	2.12	46.20
	10%	71.48	2.14	2.19	93.33

Figure 3. Per contaminant concentration measurements. NanoDrop measurements also include 260/280 and 260/230 absorbance measurements.

Sample Name	Protocol	P0	P1	P2	Average Read Length	Average Read Quality
control 0.2nM	Diffusion	17314	28275	29564	2944.35	0.829
phenol 0.2nM	Diffusion	56351	0	18802	0	0
sds 0.2nM	Diffusion	2976	0	72177	0	0
heparin 0.2nM	Diffusion	223	3	74927	1849.67	0.807
RNA 0.2nM	Diffusion	22614	9	52530	2440.78	0.810
control 0.05nM	MagBead	9836	39671	25646	2763.68	0.816
phenol 0.05nM	MagBead	17020	36225	21908	2639.08	0.830
sds 0.05nM	MagBead	71503	0	3650	0	0
heparin 0.05nM	MagBead	62466	1	12686	0	0
RNA 0.05nM	MagBead	44487	4798	25868	2605.27	0.827

Figure 4. Sequencing results for the samples that didn't sequence using the Standard Seq v2 diffusion loading protocol. Sequencing was done with both diffusion and MagBead protocols.

Acknowledgements

The authors would like to thank Wendy Weise, Kathryn Keho, Edwin Hauw, Jim Burrows, Kitland Louie, Uyen Nguyen, Trevin Rard and Deana Arnold.

Conclusions

When considering sequencing with the PacBio® *RS*, sample quality should be considered first. Contamination can be carried through from extraction chemicals, purification methods, growth media, etc. Using methods such as quantitation by both NanoDrop® and Qubit® systems, as well as gel electrophoresis, one can determine the quality of their starting DNA and possible effects on sequencing.

Accurate quantitation is critical to successful SMRT® sequencing. You should measure concentrations by using two methods of quantitation, such as the NanoDrop® and Qubit® systems. Though the Qubit® fluorometer is a more accurate measure of concentration, if concentration measurements from the two methods are grossly different, the sample is likely contaminated and will need to be purified prior to SMRTbell™ library preparation. It is also important to use a NanoDrop® spectrophotometer to gain information on absorbances and their ratios. Since DNA absorbs ultraviolet (UV) light at 260 nM, the A260/280 ratio is often used to determine purity. Since some contaminants absorb UV light in the same range of DNA, they will inflate the DNA concentration. Other contaminants, which absorb UV light in a different range, may throw off the A260/280 or the A260/230 measurements. A great deal of information on sample quality can be gained by assessing concentration, 260/280 and 260/230 measurements together. Measurements that are significantly different from a control sample should be further purified before continuing with library preparation (*Figures 1, 3*).

DNA purity can also be assessed by using gel electrophoresis. In *Figure 2* there are five strains of the same sample that were run on a 1.2% Lonza FlashGel® with a 0.1-4 kb DNA ladder. For gel analysis, samples should contain a single band of DNA. Smears, or secondary bands are often signs of contamination or degraded DNA samples. This type of DNA is not of high quality and will likely result in poor sequencing unless contaminants are removed.

Contaminants are known to affect sequencing in two ways. The first, as previously discussed, is by impacting concentration measurements. When concentration is falsely high it results in too little DNA going through the SMRTbell™ library preparation and too much polymerase being added in the polymerase binding step. Having too little DNA and too much polymerase causes low loading and high amounts of noise in the ZMWs. The second way contaminants affect sequencing is by affecting the binding of the polymerase to the templates. This will result in the same phenotype as the first. This binding competition is why some samples can't be recovered, even with MagBead station loading (*Figure 4*). Although, in some instances MagBead aids in sequencing recovery, it is most important that the sample is free of contamination and accurately quantitated before beginning SMRTbell™ library preparation.

References

[1]Thermo Fisher Scientific. "NanoDrop 1000 Spectrophotometer V3.7 User's Manual." (2008).
[2]invitrogen™. "Qubit™ dsDNA HS Assay Kits." (2010).
[3]Pacific Biosciences. "Template Preparation and Sequencing Guide." (2010).