Low-Diversity Sequencing on the Illumina MiSeq® Platform

New version of MiSeq Control Software facilitates low-diversity analysis.

Introduction

Illumina recently released a new version of MiSeq Control Software (MCS) that includes updated Real Time Analysis software (RTA), version 1.17.28. These updates significantly improve the data quality of low-diversity samples with as little as 5% PhiX DNA spike-in and no other modifications. Previously, the MiSeq platform required library prep modifications or significant PhiX spike-in to successfully sequence low-diversity samples. This technical note describes the changes in RTA 1.17.28 and demonstrates the results of Illumina's internal validation testing of low-diversity samples.

Improvements in RTA Software

The MiSeq System must be able to maintain focus, register images to the cluster map, and make proper base calls to deliver high-quality data. When sequencing low-diversity libraries, these functions can be impaired. The software changes in RTA 1.17.28 address this challenge and include the following improvements:

- **Spot-finding algorithm**: Improves template generation and allows higher sensitivity for optically dense and dark images.
- **Template matrix calculation**: Accommodates denser and more accurate templates.
- **Color matrix calculation**: Performed at the beginning of read 1 using 11 cycles of data to increase the likelihood of diversity, which is then saved for all future reads. Hard coding is no longer suggested for most cases.
- **Empirical phasing correction algorithm**: Optimizes the phasing and prephasing corrections for each cycle and tile to maximize the mean chastity of the intensity data. The result is that RTA is no longer dependent upon an accurate rate calculation, since the best correction is applied at every cycle, but instead performs cycle-by-cycle corrections that are analyzed at cycle 25. This analysis gives a calculated rate that is then saved to the InterOp file and displayed in Sequence Analysis Viewer (SAV).

Testing of Low-Diversity Samples

Illumina has completed significant internal testing to validate these changes in RTA primary analysis and establish the performance of low-diversity runs. The data presented in Table 1 directly compare the data analyzed by the previous RTA 1.17.22 (flow cells A2AFU and A0U8Y), with the same data which were re-analyzed by the current RTA 1.17.28.

These data sets demonstrate that RTA 1.17.28 can significantly improve analysis of low-diversity samples with no library prep manipulation or software changes and at least 5% PhiX spike-in. Greater amounts of PhiX do not negatively impact sequencing quality.

Table 1. Comparative data from RTA versions

Flow Cell ID	Sample Type	Run Length	RTA Version	Cluster Density k/mm^2	% PF	% Q30	% PhiX (Aligned)
A2AFU	Single Amplicon	2 x 88	1.17.22	370	19	54.2	0.1
A2AFU	Single Amplicon	2 x 88	1.17.28	576	88	97.5	5
A0U8Y	Single Amplicon	2 x 55	1.17.22	687	92	75	6
A0U8Y	Single Amplicon	2 x 55	1.17.28	762	91	93.5	5
A2GYT	16S	2 x 251	1.17.28	1020	81	82.8	7

Further Considerations

Users may notice that some quality metrics such as intensity plot and Q-score plot in low-diversity runs appear different from balanced samples with the new RTA.

- Normal cluster density (500–1200 k/mm^2) should be used with low-diversity samples. Illumina has seen that very low densities with low-diversity libraries can result in optical duplicates in the template, due to base call information not being used for the template generation logic.
- Phasing and prephasing rates in SAV will no longer be available at cycle 12. These rates are now calculated at cycle 25.
- Phasing and prephasing rates should not be used as a metric of run quality or instrument health for low-diversity samples. It can be normal for the reported phasing and/or prephasing rates to be inaccurate or anomalous, but it is important that the proper correction is being applied and data quality is not sacrificed. This is purely a reporting issue in the calculation of rates used in SAV.
- When sequencing low-diversity samples, Q scores and PhiX error rate plots displayed by SAV may appear anomalous for low-diversity runs. As explained above, small but abrupt variations in the PhiX error rate and Q score plots are normal for low-diversity runs while overall accuracy rates are maintained. Users should judge run quality on overall rates and not individual cycles. This is because the metrics are averaged across all clusters and when the clusters are identical at multiple positions the metrics are no longer population-averaged. Figure 1 shows typical SAV data from a low-diversity run.

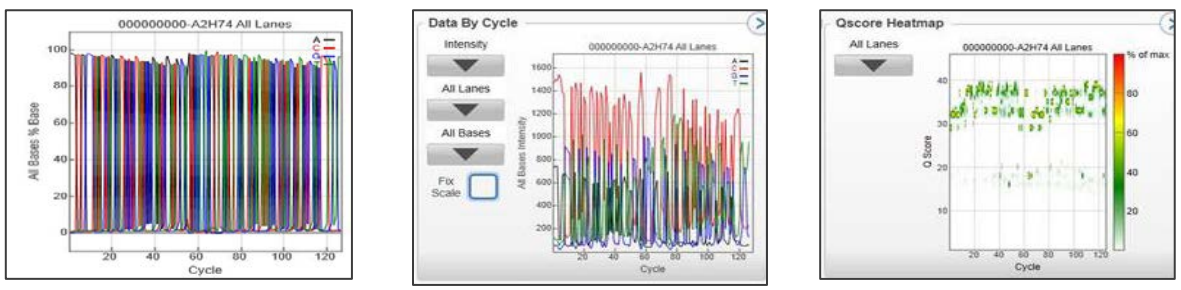
Troubleshooting Difficult Samples

While RTA primary analysis has been significantly improved to handle the low-diversity cases described in this technical note, certain sample features can prevent these changes from functioning properly, resulting in poor data quality.

- No PhiX and no A or C in the first four cycles: If the sample contains only T and C in the first four cycles, image registration will fail.

Figure 1: SAV Data from Low-Diversity Samples

Percent Base **Intensity** **Q-Score Plot**



Examples of standard quality metrics generated from an in-house run.

- Observance of all four bases in the first 11 cycles of the run: If one or more bases are not present in the first 11 bases of the sequence (such as in a long homopolymer repeat), then matrix calculation will be inaccurate and run quality is likely to be impacted. This is because the color matrix calculation requires the presence of all four bases during the first 11 cycles.

Summary

Using RTA 1.17.28, Illumina has successfully performed very low diversity runs on the MiSeq: mostly single-amplicon and 16S samples with varying run lengths. The quality of these experiments was validated by alignment and variant calling of known samples. Runs that returned very poor data quality with earlier versions of RTA, deliver high-quality results with the new version of the software. The best results were achieved with a PhiX spike-in of at least 5% to improve both software robustness and sequencing accuracy.

Remaining limitations of the analysis pipeline have been detailed in this technical note. Excluding these run conditions, Illumina has demonstrated that the MiSeq System can support the sequencing of very low-diversity samples, without hard coding RTA parameters or using sophisticated library preparation when running samples at normal cluster density and with at least 5% PhiX spike-in.

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

Pub. No. 770-2013-013 Current as of 20 November 2014