VeriSeq PGS-MiSeq QC Assessment Guide

This technical note provides guidance on the quality control metrics required for the VeriSeq PGS-MiSeq application.

Introduction

VeriSeq PGS-MiSeq is an application that uses next-generation sequencing (NGS) on a MiSeq to provide pre-implantation genetic screening (PGS) of all 24 chromosomes for aneuploidy. The test can be used on a single cell or few cells from an embryo biopsy. VeriSeq PGS is a research use only (RUO) product. This document provides guidance in the interpretation of sequencing data and results achieved as part of the test. It is intended to assist with the assessment of whether a sequencing run is successful and whether a sample shows quality control (QC) values that are considered acceptable. In addition, recommended dsDNA concentrations following SurePlex amplification are provided in the summary table at the end of the document.

Refer to the VeriSeq PGS Library Preparation Guide (part # 15052877) and the BlueFuse Multi Reference Guide (part # 15053620) for more information on how to perform the test and the analysis of the results.

Refer to the MiSeq System User Guide (part # 15027617) for run metrics of the instrument. For more information on the MiSeq, see the online training video[1] on the Illumina support site.

Library Handling and Sequencing Issues

BlueFuse Multi v4.1 or later can help to assess the quality of VeriSeq PGS-MiSeq data on two different levels – flow cell level or sample level. The Sequencing QC Report lists a number of run-level QC values. If the Summary.xml file from the MiSeq output folder is provided during the import or automatically found in the run folder during batch import, this report is generated.

To find out more about the issues identified or to monitor and analyze any other possible issues, use Illumina's Sequence Analysis Viewer (SAV)[2] or BaseSpace[3] (if activated for the run). The SAV can read sequencing QC information from a MiSeq run folder and display key metrics to troubleshoot the most common sequencing problems.

Figure 1: Part of the Sequencing QC Report from BlueFuse Multi providing run-level QC data

Sequencing QC Measures	
Sequencing Lane	
Lane ID:	1
Aligned Reads (PF) %:	78.99
Sequencing Run	
Cluster Number:	34770888
Cluster Number Passing Filter:	28224083

- **Sequencing Lane–**Because the MiSeq flow cells currently contain a single lane, only information for lane 1 is listed here. Some of the recorded reads do not reliably match to the human genome that is used as a reference. The percent of ***Aligned Reads (PF)*** is reported and is ideally between 80 to 85% or at least greater than 70% of the total number of reads. Lower values could indicate contamination in your DNA source, biopsy quality, or suboptimal whole genome amplification.
- **Sequencing Run–**The total ***Cluster Number*** corresponds to how many positions with distinct DNA molecules are detected on the flow cell surface (raw clusters). The number is ideally close to 30 Million (M), and reaches at least 25 M raw clusters. Less than 25 M or more than 30 M clusters indicate that the amount of DNA added to the flow cell requires optimization for subsequent sequencing runs. While lower numbers do not provide enough output for a reliable analysis, higher numbers result in misreading of the sequence. The SAV can provide more detailed information about the cluster density and should be used after every run. Optimal density is between 1,200k and 1,400k/mm2.
  From all the raw clusters that formed on the flow cell, only a subset is considered to be reliable enough to be used in downstream processing. Ideally this subset, the "passing filter" (PF) clusters, reaches at least 85% of the total number of raw clusters. If lower values of PF clusters are reached (i.e. 75%), it indicates that the amount of denatured library used for seeding the flow cell was suboptimal. In absolute numbers, this result would relate to a minimum of around 19 M clusters reported as ***Cluster Number Passing Filter.***
  The ***Yield (kb)*** is a result of the number of PF clusters and the read length (36 base pair) and is ideally around 900,000 kb.

Sample Issues

The sample-level report of BlueFuse Multi assesses the quality of the data from a specific sequenced sample as it is processed from initial base calling, to alignment, filtering, and copy-number calling.

Figure 2: Sample-level QC values in BlueFuse Multi used to access quality of data for individual samples

Sample QC measures	
QC Status:	Pass
Average quality score:	35.6
Average alignment score:	35.4
Number of total reads:	1237354
Number of mapped reads:	1017145
Number of reads after filtering:	808646
Overall noise (DLR):	0.24

- **QC Status–**The overall QC status of the sample, by default, is listed as Pass. Alternatively, the QC status can be set as QC Fail. Do not store in history or Suboptimal Pass using the Sign Off function.
- **Average Quality Score–**The base quality score (Q-score) is a measure of confidence in the base that was called at a given position. The Q-score is calculated as mean(mean(phred33-based base quality)per read) for all reads per sample. A Q-score of 30 estimates that there are 0.1% wrong bases in the result. For the report, this score is recalculated after filtering and is ideally around 35 with a minimum score of 30.
- **Average Alignment Score–**The alignment software calculates this score and takes into account the Q-score, read length, and the number of alignments for every read. The average alignment score is ideally around 35 with a minimum score of 30.
- **Number of Total Reads–**For a well-balanced library with 24 samples, there is ideally around 1 M reads reported for each sample, with a minimum of 700,000 reads
- **Number of Mapped Reads–**From the total number of reads, 80% ideally align to the genome, i.e. around 800,000. Lower values can indicate a contamination in your DNA source or suboptimal whole genome amplification.
- **Number of Reads after Filtering–**Aligned reads are further filtered. The number of filtered reads usable for copy-number calling is ideally around 60% of the total reads, i.e. 600,000 reads. The minimum value is 250,000 reads, but samples with this number of reads can result in noisy profiles that are difficult to interpret.
- **Overall Noise (DLR-Derivative Log Ratio)–**The overall sample noise measures the spread of the difference in copy number values between all bins of a chromosome. Values are ideally less than 0.3. Values above 0.4 indicate low quality sample DNA or problems during the amplification steps.

Table 1: Summary of VeriSeq PGS-MiSeq Acceptance Criteria

	Parameter	Optimum Value	Acceptance Value	Source of Values
Amplification Test	Gel Electrophoresis	Positive	Positive	Laboratory tests
	Concentration dsDNA ng/µl (single cells)	25 – 35	> 10	
Flow Cell-based Parameter	Density (K/mm^2)	1200 –1400	1100 – 1600	"Run summary" page on BaseSpace or "Summary" in SAV
	% of clusters passing filter (PF)	85	75	
	Total number of reads	30,000,000	25,000,000	
	Total number of reads (PF)	25,000,000	19,000,000	
	% of reads reaching Q30	95	90	
	Maximum phasing/pre-phasing	0.5 / 0.05	0.65 / 0.1	
	% of reads identified (PF)	95	85	"Indexing QC" page on BaseSpace or "Indexing" in SAV
Sample-based Parameter	% of reads identified (PF) per sample	5	2.5 – 8	
	% of reads identified (PF) for empty wells	< 0.025	< 0.05	
	Number of total reads	1,000,000	700,000	Downstream analysis from BlueFuse Multi
	Number of reads after filtering	500,000	250,000	
	% of total reads after filtering	> 50	> 35	
	Average quality score (Q-score)	> 35	> 30	
	Average alignment score	> 35	> 30	
	Sample noise score (DLR)	0.2	< 0.4	
Chromosome-based scores	Region Confidence (value in CNV Table)	1.0	> 0.7	Downstream copy-number analysis from BlueFuse Multi

References

[1]http://support.illumina.com/training/courses/MiSeq_Sequencing_Fundamentals/

[2]https://basespace.illumina.com

[3]https://support.illumina.com/sequencing/sequencing_software/sequencing_analysis_viewer_sav.ilmn

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

FOR RESEARCH USE ONLY

© 2014 Illumina, Inc. All rights reserved.
Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPro, CytoChip, DesignStudio, Epicentre, GAIIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, ForenSeq, MiSeq, MiSeqDx, MiSeq FGx, NeoPrep, Nextera, NextBio, NextSeq, Powered by Illumina, SeqMonitor, SureMDA, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode, verifi, VeriSeq, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub. No. 770-2014-030 Current as of 23 July 2014

illumina®