Performance of NEXTFLEX® Variant-Seq™ SARS-CoV-2 Kit using COVID-19 Positive Samples

APPLICATION NOTE

NEXTFLEX® Variant-Seq™ SARS-CoV-2 Kit

The NEXTFLEX® Variant-Seq™ SARS-CoV-2 Kit identifies all mutations in a SARS-CoV-2 sample using Next Generation Sequencing (NGS).

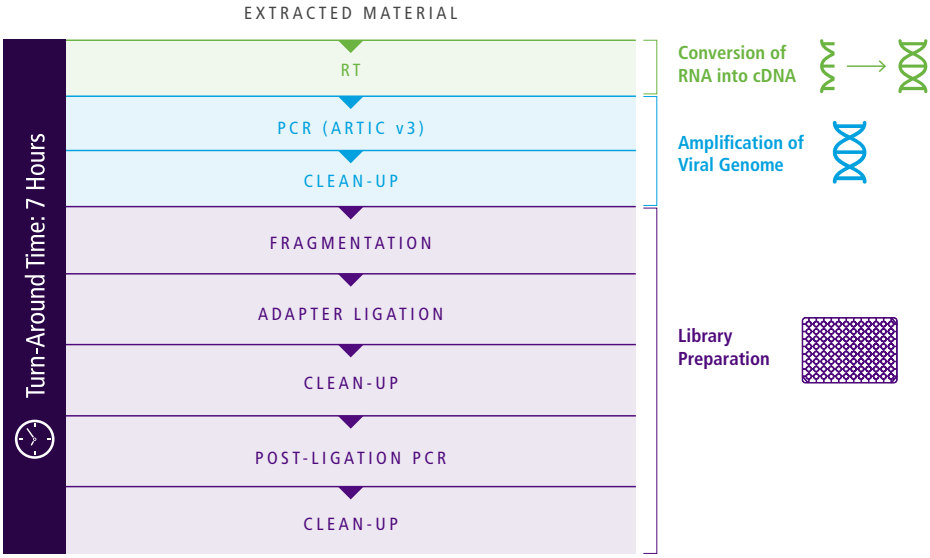
Introduction

The NEXTFLEX® Variant-Seq™ SARS-CoV-2 Kit identifies all mutations in a SARS-CoV-2 sample using Next Generation Sequencing (NGS). This is a powerful method to understand the full extent of diversity of variant SARS-CoV-2 viruses circulating in a population. The genomic data obtained provides information on multiple aspects that are relevant to understand and manage the global pandemic, such as early detection and characterization of emerging variants, virus transmission dynamics and the impact of response measures on the spread of the virus in a population.

One of the challenges of variant identification confidence is the variability in the Ct value of PCR-positive samples. The Ct value has been linked to the number of viral genome copies present in a sample, raising the question of how well the NGS workflow performs across different Ct values. We address this question in this note with data obtained using clinical samples.

The NEXTFLEX® Variant-Seq™ SARS-CoV-2 kit has been optimized for use with 1 million clusters per sample at 1x36bp. We also address in this note the question of the performance the NEXTFLEX® Variant-Seq™ SARS-CoV-2 kit using different read length and read depths, with COVID-19 positive samples.

NEXTFLEX® Variant-Seq™ SARS-CoV-2 workflow



PerkinElmer®
For the Better

Method

A sample set of 636 nucleic acid extracts that tested positive for SARS-CoV-2 with the PerkinElmer® SARS-CoV-2 Nucleic Acid Detection Kit were selected. The Ct value for the N Gene of these samples ranged from 9.21 up to 32.91 (Figure 1).

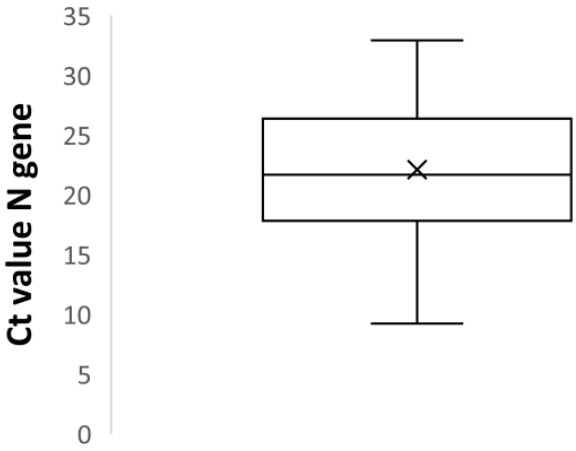


Figure 1. Boxplot showing distribution of samples according to Ct value for the N Gene

An 8 µL aliquot of total extracted nucleic acids were used as input for the NEXTFLEX® Variant-Seq™ SARS-CoV-2 kit for all samples, regardless of their Ct value. The final libraries were quantified using the Thermo Fisher® Scientific Qubit™ HS dsDNA kit and then run on the LabChip® GX Touch™ for fragment size analysis. After pooling, the concentration of the pool was 6.72 ng/µL and fragment size averaged 434 bp. Sequencing was completed on an Illumina® NovaSeq® 6000 instrument at 1x36bp. To study the effect of different sequencing conditions, a subset of samples was re-sequenced on an Illumina® MiSeq® instrument at 1x36, 1x75, 2x36 and 2x75 bp.

FastQ files were uploaded to the CosmosID® SARS-CoV-2 Strain Typing Analysis Portal for analysis. Data was also run through the Illumina® DRAGEN COVID Lineage App for comparison. SARS-CoV-2 genome coverage was also reviewed with the Integrative Genomics Viewer software (IGV).

Results

Relationship Between Ct Value & % of SARS-CoV-2 Genome Bases Covered

To understand the performance of the NEXTFLEX® Variant-Seq™ SARS-CoV-2 kit workflow on different viral loads, we first plotted the percentage of bases of the genome covered with a coverage depth of at least 10x, against the Ct value (Figure 2).

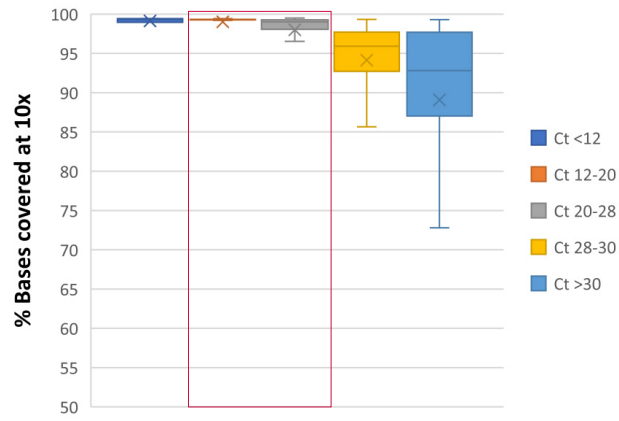


Figure 2. Box plot representing % of bases of the viral genome covered against Ct value for the N gene. The red box denotes the recommended Ct value input range for the NEXTFLEX® Variant-Seq™ SARS-CoV-2 kit.

The average percentage of the SARS-CoV-2 genome represented in the sequencing data ranged from >99% when Ct values were less than 28 to >92% for very high Ct values above 30. The correlation coefficient r= -0.428 for the entire data set. However, if we focus on samples with Ct values between 12-28 then the r= -0.247, indicating that in this recommended range, the percentage of bases covered is relatively independent of Ct value, and remains above 99%.

Relationship Between Ct Value & Average Base Coverage

Samples with read counts above 1,000,000 were downsampled for this analysis. We plotted the average base coverage of the SARS-CoV-2 genome against the Ct value (Figure 3). In this case, we found a stronger negative correlation between the parameters, with r=-0.7408.

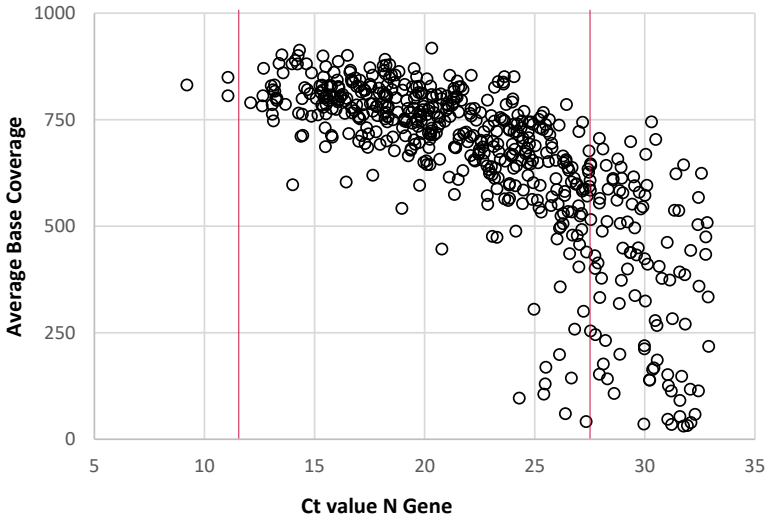


Figure 3. Average Base Coverage against Ct value of the N gene. The red lines mark the lower and upper limit of the recommended Ct range for this kit.

These data suggest a correlation between increasing Ct value and decreasing genome depth and breadth of coverage.

Relationship Between Ct Value & Result QC

Using our recommended minimum 91% of bases covered at 10x depth to allow for confident variant calling, there were more samples that did not reach the acceptance criteria threshold when the Ct value was above 28, than when it was below 28 (Figure 4).

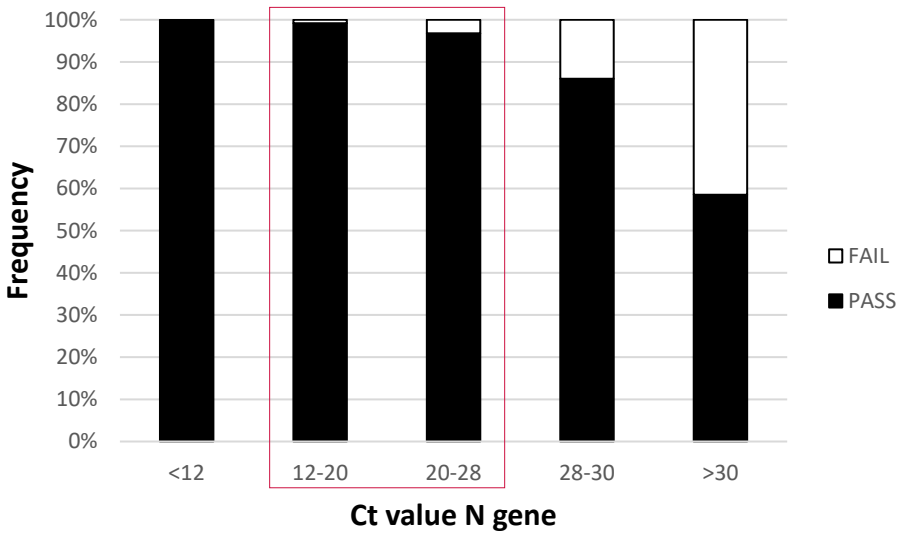


Figure 4. Classification of the results as PASS or FAIL according to the QC described in the text. The red box marks the recommended Ct range for the kit.

Relationship Between Ct Value & Mutations Detected

We quantified the number mutations detected for each of the samples that were classified as PASS for variant calling (n=628) and investigated whether there was any difference in the number of average mutations detected in different Ct categories. We found that there were no differences between the different Ct groups, suggesting that there is no link between Ct and number of variants.

Table 1. Average number of mutations detected per sample within each Ct category

Ct value of N gene	Average number of mutations
<12	20 (7 samples)
12-20	21 (247 samples)
20-28	21 (279 samples)
28-30	19 (49 samples)
>30	20 (46 samples)

Concordance between CosmosID® SARS-CoV-2 Strain Typing Analysis & Illumina® DRAGEN COVID Lineage App

Lineage was independently assigned by the CosmosID® SARS-CoV-2 Strain Typing Analysis software and with Illumina® DRAGEN™ COVID Lineage App. We confirmed 100% concordance in the lineages assigned by both applications, confirming that both pipelines are equivalent.

Impact of Read Depth on Performance

Two samples (A and B) were selected and were run on an Illumina® MiSeq® instrument at 1x36bp to analyze the effect of number of reads on data quality and results. The data was downsampled to 1M, 750K, 500K and 300K reads and uploaded to the CosmosID® SARS-CoV-2 Strain Typing Analysis Portal for calculation of metrics and lineage assignment. Results are listed in Table 2.

Table 2. Performance of samples A and B at different read depths.

Sample ID	Number of reads analyzed	% Bases Covered at depth ≥10x	Average Base Coverage	Pangolin lineage
Sample A (Ct=11.95)	1,000,000	99.4	845.213	B.1.402
	750,000	99.37	633.827	B.1.402
	500,000	99.17	422.456	B.1.402
	300,000	99.05	253.425	B.1.402
Sample B (Ct=32.04)	1,000,000	97.5	321.229	B.1.240
	750,000	96.9	240.523	B.1.240
	500,000	95.52	160.191	B.1.240
	300,000	92.29	96.1313	B.1.240

All read counts per sample passed the QC threshold of at least 91% of bases covered with depth ≥ 10x. As expected, the Pangolin lineage assigned to each sample was not affected by the read depth. The average base coverage decreased in a linear proportion with the number of reads included in the analysis (r=0.999). It is noteworthy that Sample B has a Ct higher than the upper limit recommended by the NEXTFLEX® Variant-Seq™ SARS-CoV-2 kit.

Impact of Read Depth on Performance

Shorter read lengths are more cost-effective to sequence and more efficient in terms of analysis time. However, they can be more difficult to map. In theory, longer reads should lead to higher percentage of mapped reads and therefore higher coverage. To analyze the impact of read length, we re-sequenced samples A and B and compared the results with those obtained at 1x36 bp (Table 3).

Table 3. Percentage of mapped reads and average base coverage of Sample A and B at different read lengths.

Sample ID	Number of reads analyzed	1x75 bp		1x36 bp	
		% of Mapped Reads	Average Base Coverage	% of Mapped Reads	Average Base Coverage
Sample A (Ct=11.95)	1,000,000	96.12	2124.85	70.52	845.213
	750,000	96.12	1593.69	70.51	633.827
	500,000	96.15	1062.85	70.49	422.456
Sample B (Ct=32.04)	1,000,000	94.56	1087.23	26.78	321.229
	750,000	92.81	815.077	26.74	240.523
	500,000	89.91	543.653	26.71	160.191

As expected, the percentage of mapped reads and average base coverage was higher at 1x75 than at 1x36. We noticed that doubling the read length produced an average increase of 2.95-fold in the coverage at all read depths.

Impact of Paired-end Sequencing on Performance

To analyze the impact of using a paired-end sequencing mode on the performance of the NEXTFLEX® Variant-Seq™ SARS-CoV-2 workflow, we re-sequenced samples A and B at 2x75 and 2x36 bp. The effect of paired-end mode compared to single-end is to increase the number of reads available for analysis.

To verify this, we checked the results of both samples A and B at 2x75 bp using a cluster depth of 500,000 (1,000,000 reads). The results in Table 4, show the average base coverage is equivalent with single read and paired-end sequencing.

Table 4. Results obtained when samples were run at 2x75 bp and 500,000 cluster depth compared to the sample samples run at 1x75 bp and 1,000,000 cluster depth.

Sample ID	Sequencing Mode (bp)	Cluster Depth	Number of reads analyzed	% Bases Covered at depth >10x	Average Base Coverage
Sample A	2 x 75	500,000	1,000,000	99.95	2148.86
(Ct=11.95)	1 x 75	1,000,000	1,000,000	99.94	2124.85
Sample B	2 x 75	500,000	1,000,000	98.81	1113.15
(Ct=32.04)	1 x 75	1,000,000	1,000,000	98.88	1087.23

Conclusions

The NEXTFLEX® Variant-Seq™ SARS-CoV-2 workflow can sequence the SARS-CoV-2 genome from positive COVID-19 samples across a wide range of Ct values. Coverage and rates of success are highest when the Ct value of the sample is below 28, as recommended in the kit manual.

For a given Ct range, coverage is proportional to read length and depth. Increasing read length from the recommended 36 to 75 bp would allow users to obtain high quality results across a broader range of Ct values.

PerkinElmer, Inc.
940 Winter Street
Waltham, MA 02451 USA
P: (800) 762-4000 or
(+1) 203-925-4602
www.perkinelmer.com

PerkinElmer®

AG012105_07_AP PKI