

Validation of Clinical Quality Control Metrics for SARS-CoV-2 Sequencing Using the Ion Torrent Genexus Integrated Sequencer

TALBOT D. LEATHERS, MARTINA ZAMPONI, MACKENZIE TARDIF-KUNK, OYUKY GARCIA, KASEY N. PARKER, MEGAN A. HEPT, SHAMIR HINES-BATTLE, AARON HARRISON, TABETHA R. SUNDIN, PETER A. MOLLIKA

ABSTRACT

Since its emergence in 2019, SARS-CoV-2 has infected millions globally, with novel variants arising through mutations that enhance virulence or immune evasion. Given the shortage of qualified laboratory professionals during this time, clinical platforms exhibiting minimal hands-on time gained popularity. One such platform, Thermo Fisher Scientific's Ion Torrent Genexus, provided sequencing capabilities to short-staffed laboratories, yet standard quality control metrics for this platform remain limited. In this study, we established quality metrics for accurate variant calling on the Genexus Integrated Sequencer. Upper respiratory

specimens positive for SARS-CoV-2 were sequenced and rigorously assessed using the proposed metrics. We sequenced 339 patient specimens and 89 controls across varying viral loads. Among positive controls that met the metrics, 100% were assigned correct lineages, while 86% of negative controls appropriately failed quality thresholds. The proposed metrics provide a framework for accurate SARS-CoV-2 variant identification on the Genexus platform, offering a model for clinical laboratories adopting this technology for reliable sequencing and variant tracking.

ABBREVIATIONS: CDC - Centers for Disease Control and Prevention, CLIA - Clinical Laboratory Improvement

Talbot D. Leathers, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia; Molecular Diagnostics, Sentara Norfolk General Hospital, Norfolk, Virginia

Martina Zamponi, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia; Molecular Diagnostics, Sentara Norfolk General Hospital, Norfolk, Virginia; Macon & Joan Brock Virginia Health Sciences Eastern Virginia Medical School at Old Dominion University, Norfolk, Virginia

Mackenzie Tardif-Kunk, Department of Medical Diagnostic and Translational Sciences, Ellmer College of Health Sciences, Old Dominion University, Norfolk, Virginia

Oyuky Garcia, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia; Core Laboratory, Riverside Regional Medical Center, Newport News, Virginia

Kasey N. Parker, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia; Department of Biological Sciences, Old Dominion University, Norfolk, Virginia

Megan A. Hept, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia; Department of Chemistry, Old Dominion University, Norfolk, Virginia

Shamir Hines-Battle, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia

Aaron Harrison, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia; Molecular Diagnostics, Sentara Norfolk General Hospital, Norfolk, Virginia

Tabetha R. Sundin, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia

Peter A. Mollica, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia; Molecular Diagnostics, Sentara Norfolk General Hospital, Norfolk, Virginia; Department of Medical Diagnostic and Translational Sciences, Ellmer College of Health Sciences, Old Dominion University, Norfolk, Virginia

Address for Correspondence: Peter A. Mollica, Monarch Molecular Clinical Laboratory, Old Dominion University, Norfolk, Virginia; Molecular Diagnostics, Sentara Norfolk General Hospital, Norfolk, Virginia; Department of Medical Diagnostic and Translational Sciences, Ellmer College of Health Sciences, Old Dominion University, Norfolk, Virginia, pmollica@odu.edu

Amendments, CRM - commercial reference material, Ct - cycle threshold, LOD - limit of detection, NGS - next-generation sequencing, NTC - no template control, PCR - polymerase chain reaction, QC - quality control, qPCR - quantitative PCR, VTM - viral transport media.

INDEX TERMS: SARS-CoV-2, NGS, QC metrics, Ion Torrent Genexus Integrated Sequencer, amplicon-based sequencing.

Clin Lab Sci 2025;38(1):10–18

INTRODUCTION

SARS-CoV-2 is a positive-sense, single-stranded RNA virus that first emerged in Wuhan, China, in late 2019. This member of the *Coronaviridae* family is highly infectious and causes a range of respiratory symptoms in humans, from asymptomatic cases to severe illness and death. The high mutability of SARS-CoV-2 within its 30-kb genome has contributed significantly to the virus's rapid and widespread transmission. Novel mutations, particularly those affecting the spike protein, have increased the virus's virulence and capacity for immune evasion, resulting in distinct SARS-CoV-2 variants.¹⁻³ Identifying the prevalence of these variants within a population is essential both during and before outbreaks, as it helps inform appropriate public health measures to prevent further disease spread.^{4,5}

Throughout the pandemic, a variety of next-generation sequencing (NGS) methods and assays have become available to clinical laboratories for SARS-CoV-2 variant analysis.^{6,7} Many of these assays were adapted to preexisting sequencing platforms.⁷⁻⁹ Among these, Illumina's technology has been favored because of its reliability and extensive bioinformatics support, which has led public health departments, hospital systems, and private laboratories to invest in Illumina instrumentation long before the SARS-CoV-2 pandemic.^{10,11} However, these platforms require significant hands-on time from skilled laboratory scientists, a drawback when rapid, high-volume sequencing is needed, as seen during the COVID-19 pandemic.¹²⁻¹⁵

Selecting an NGS platform and assay requires consideration of factors such as reagent availability, cost, maintenance needs, expected volume, turnaround time, and laboratory scientist's hands-on time.^{13,16-18} Thermo Fisher's Ion Torrent Genexus Integrated Sequencer offers an alternative solution by requiring minimal hands-on time through automated library preparation, sequencing, and onboard analysis, a critical advantage amidst a national shortage of skilled laboratory personnel.^{12-15,19-21} Although Thermo Fisher has expanded its clinical support infrastructure to facilitate SARS-CoV-2 sequencing, there remains a lack of uniform quality control (QC) guidelines specific to clinical reporting of SARS-CoV-2 variants on the Genexus platform.^{7-11,13,22}

The SARS-CoV-2 Insight Panel on the Genexus platform provides full-genome sequencing of SARS-CoV-2 via amplicon-based targeted NGS.⁶ This process takes approximately 30 hours, during which previously extracted RNA undergoes library preparation, templating, sequencing, and analysis through the Ion Torrent Software.¹⁷ Downloadable data files are generated from each run, facilitating further analysis and troubleshooting.²³ Although the software provides several metrics related to both run and sample quality, no guidelines have been established by the manufacturer to assess data quality. This was due in part to the novelty of the SARS-CoV-2 virus and the research use designation of this assay. This lack of standardization poses a risk for inaccurate results, as incomplete data can lead to false-positive results or incorrect lineage assignments if not rigorously vetted. Prior studies on the Genexus platform have validated its ability to reliably generate complete SARS-CoV-2 genomes.^{6,23}

Despite the widespread availability of SARS-CoV-2 sequencing services in clinical and research laboratories, there remains a gap in uniform quality metrics for sequence quality determination. Although the Ion Torrent Genexus Integrated Sequencer shows promise for clinical applications, it was originally designed for research purposes and lacks a well-defined clinical support workflow to meet the needs of clinical laboratories.^{7-11,13,22} While existing quality metrics for NGS platforms primarily apply to other instruments, the limited metrics currently available for the Genexus platform pertain only to individual sequence quality^{8,12,24} and do not address the increased stringency for in vitro diagnostic testing. Establishing stringent, standardized clinical metrics is essential to maintain a high standard of patient care in laboratory reporting. This article addresses this need by proposing a set of robust quality metrics for clinical use to evaluate run and sample quality for SARS-CoV-2 sequencing on the Genexus Integrated Sequencer.

MATERIALS AND METHODS

Patients and Samples

Nasopharyngeal and midturbinate swabs were collected in sterile viral transport media (VTM) collection devices. All samples were transported to the laboratory under cold conditions (2–6°C) within 6 hours of collection. COVID-19 diagnostic status for each specimen was determined in a Clinical Laboratory Improvement Amendments (CLIA)-certified, high-complexity clinical laboratory using Food and Drug Administration Emergency Use Authorization–approved methods, primarily the Cobas 6800 SARS-CoV-2 Kit, Diasorin Simplexa COVID-19 Direct Kit, or BioFire Respiratory 2.1 (RP2.1) Panel. A complete workflow for these specimens is provided in Figure 1.

RNA Extraction and Quantification

RNA isolation was performed in a Biosafety Level 2 plus pre-amplification environment. Total nucleic acid extraction was

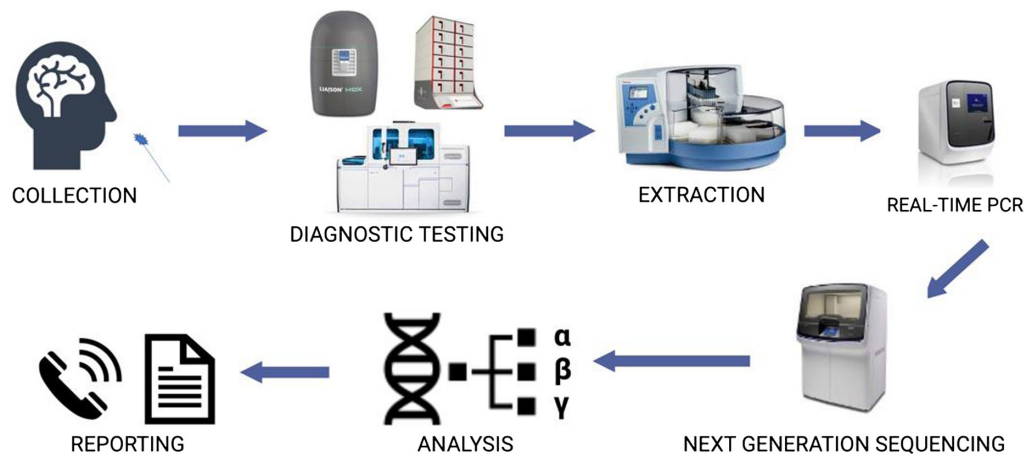


Figure 1. SARS-CoV-2 specimen handling and processing workflow. Nasopharyngeal and midturbinate specimens were collected from the university's student health center and brought to the clinical molecular laboratory. After diagnostic testing, SARS-CoV-2–positive specimens were extracted and quantified for viral copy number before sequencing, analysis, and reporting. Typical turnaround time for diagnostic results was less than 1 day, allowing for sequencing of positive specimens to take place the same week, with variant prevalence reported at the week's end.

conducted on SARS-CoV-2–positive specimens using the MagMAX Viral/Pathogen II (MVP II) Nucleic Acid Isolation Kit (Thermo Fisher, category no. A48383) on a KingFisher FLEX Purification System (Thermo Fisher Scientific), following the manufacturer's instructions. The extraction protocol for 200- μ L sample input with a 50- μ L elution volume and 2 wash steps (MVP_2Wash_200_Flex.bd2) was used. For viral RNA quantification, the TaqPath COVID-19 Combo Kit (Thermo Fisher Scientific, category no. A47814) was used on a QuantStudio 7 Flex Real-Time Polymerase Chain Reaction (PCR) System (Thermo Fisher Scientific) with a 384-well block, adhering to manufacturer recommendations.²⁵ Thermocycling conditions were as follows: 25°C for 2 minutes, 53°C for 10 minutes, and 95°C for 2 minutes, followed by 40 cycles of 95°C for 3 seconds and 60°C for 30 seconds. The cycle threshold (Ct) value of the N gene from the PCR reaction–informed specimen dilution requirements for NGS library preparation. Specimens with a Ct value below 18 were diluted at least 1:10 with MagMAX Elution Buffer prior to loading onto the Genexus, per initial guidance from the Ion AmpliSeq SARS-CoV-2 Research Panel Quick Reference.^{26,27} For specimens with an N gene Ct value above 31, indicating insufficient viral copy numbers for the NGS workflow, testing was conducted to challenge the sensitivity limits of the “High Titer” and “Low Titer” assays on the SARS-CoV-2 Insight Panel. In this context, the “High Titer” and “Low Titer” assays correspond to settings within the Ion Torrent Software–labeled SARS-CoV-2 Research Assay and SARS-CoV-2 Low Titer Research Assay, respectively. Both assays use the same reagents and primer panels; however, the low-titer setting includes additional amplification cycles to enhance sensitivity for samples with lower viral loads. The Qubit RNA High Sensitivity Assay on the Qubit 4 Fluorometer was also tested as an alternative quantification method for SARS-CoV-2 NGS, in parallel with quantitative

PCR (qPCR). Negative patient specimens and sterile (uninoculated) VTM served as no-template controls. Negative patient specimens were tested as negative controls to determine the assay's ability to correctly identify the absence of the target analyte (COVID virus) as well as to conserve VTM for patient testing. Extracted RNA was stored on ice if sequencing was performed the same day or at -80°C if delayed for up to 10 days.

Library Preparation and Sequencing

Targeted NGS was conducted on extracted RNA using the Ion AmpliSeq SARS-CoV-2 Insight Research Panel (Thermo Fisher Scientific, category no. A51307) and the SARS-CoV-2 Research Panel (Thermo Fisher Scientific). The SARS-CoV-2 Insight Panel replaced the initial SARS-CoV-2 Research Panel because of its increased tolerance for emerging variants and improved performance with low-titer samples. Both panels employ 237 amplicons to cover over 99% of the SARS-CoV-2 genome, producing essentially whole-genome sequences.^{6,12} Assay selection was guided by viral copy determination, as outlined in respective panel reference materials.²⁵⁻²⁷ For specimens with N gene Ct values exceeding recommended limits, the assays were intentionally loaded to assess their performance under challenging conditions. Sequencing was conducted on both freshly extracted and previously frozen specimens. The Genexus system's high- and low-titer assays share the same primer pools but differ in amplification cycles. Manufacturer-provided controls, including the AccuPlex SARS-CoV-2 Verification Panel – Full Genome (category no. 0505-0168), Twist Bioscience SARS-CoV-2 Controls (category no. 102019, 103907, 103909, 105204), and ZeptoMetrix (category no. NATSARS(COV2)-VP), were used to assess the limit of detection (LOD), reagent lot changes,

and sequencing accuracy. These controls included Alpha, Beta, Gamma, Delta, and Omicron variants, with undiluted Zeptometrix controls at 50 000 copies/mL and Twist and AccuPlex controls diluted to concentrations ranging from 10 000 to 62.5 copies/mL. Dilutions were performed using uninoculated VTM prior to extraction and quantification. No template control (NTC) samples, originating from diagnostic testing, were processed identically to positive specimens and controls to maintain testing uniformity.

Analysis and Determination of QC Metrics and Thresholds

The Ion Torrent Reporting Software Version 6.2.1 generates a run metrics file and sample-specific files, which were applied to the metrics listed in Table 1 under the workflows shown in Figure 2. Although metric categories are defined by the system, the manufacturer does not specify cutoff criteria for data quality assessment. For each run, a run report was downloaded and reviewed following the process outlined in Figure 2A. Each run was classified as either “Passed” or “Failed.” Individual sample quality metrics (Iterative Refinement Meta-Assembler FASTA and coverage analysis summary) were reviewed (Figure 2B); samples meeting quality criteria were uploaded to PANGOLIN Version 6.04 for lineage assignment. Samples that did not produce a FASTA file or failed quality metrics lacked lineage assignment. For samples flagged as “Needs Review,” qualitative assessments included coverage overview, depth of coverage chart, Amplicon coverage chart, and read length histogram. Specifically, coverage overview indicated sequence completeness and gaps, Depth of coverage chart should show a distribution centered around 10 000 reads, Amplicon coverage chart should

display partial genome coverage, and read length histogram should have a normal distribution around 210.

For run performance, Table 1 metrics were used. While metric categories are defined, data quality thresholds are not explicitly provided by the manufacturer. These criteria and cutoff values were established and validated in our laboratory to ensure robust data generation. Following run quality assessment, individual sample metrics (Table 1) were selected for their relevance to data quality. Definitions for run and sample metrics are available in the Genexus user manual and Ion Torrent documentation.^{13,25-27} All sequences were uploaded to PANGOLIN for lineage designation using the FASTA files generated by Ion Torrent Software and mapped to the SARS-CoV-2 reference genome using Geneious Prime to assess potential coverage gaps (Supplementary Figure 1).

To further assess the reliability of the variant calls, we considered underlying sequence quality metrics beyond total read count and genome coverage. The Ion Torrent Genexus platform calculates base quality and mapping quality scores using a modified Phred scale during base calling and alignment. However, in the default SARS-CoV-2 Insight Panel workflow, these values are not readily accessible through the standard user-facing reports. While they are embedded within the raw BAM and FASTQ files, they were not directly incorporated into our QC thresholds. Instead, quality determinations were based on manufacturer-recommended and laboratory-validated summary metrics—such as mapped reads, coverage depth, uniformity, and sequence completeness. For samples marked as “Needs Review,” visual inspection of consensus sequence alignments was performed using Geneious Prime to evaluate coverage gaps and mapping consistency relative to the SARS-CoV-2 reference genome. Future refinements of this

Table 1. SARS-CoV-2 sequencing clinical quality metrics for use on the Ion Torrent Genexus

Quality Metrics		Failure Cutoff	Needs Review Cutoff	Pass Cutoff
Run metrics	Loading (%)	<80	N/A	≥80
	Enrichment (%)	<95		≥95
	Library reads (%)	<80		≥80
	Final reads (%)	<30		≥30
	Final reads/lane	<8 000 000		≥8 000 000
	Mean raw accuracy (%)	<98		≥98
Sample metrics	Total reads	<100 000	100 000–400 000	>400 000
	Mean read length	<155	≥155–174	≥175
	Uniformity of base coverage (%)	<80	≥80–<90	≥90
	Average base coverage depth	<2000	N/A	≥2000
	Percent reads on target (%)	<90	≥90–<95	≥95
	Number of mapped reads	<50 000	N/A	≥50 000
	Target base coverage at 20× (%)	<80	≥80–<90	≥90

The metrics for quality control of SARS-CoV-2 sequencing on the Genexus can be divided into run-specific metrics and sample-oriented metrics. Most run metrics listed have hard cutoffs, while the sample metrics have a review window. Sequence data that “Pass” but need review should take all listed metrics into consideration and reference the qualitative metrics prior to reporting. NA, not available.

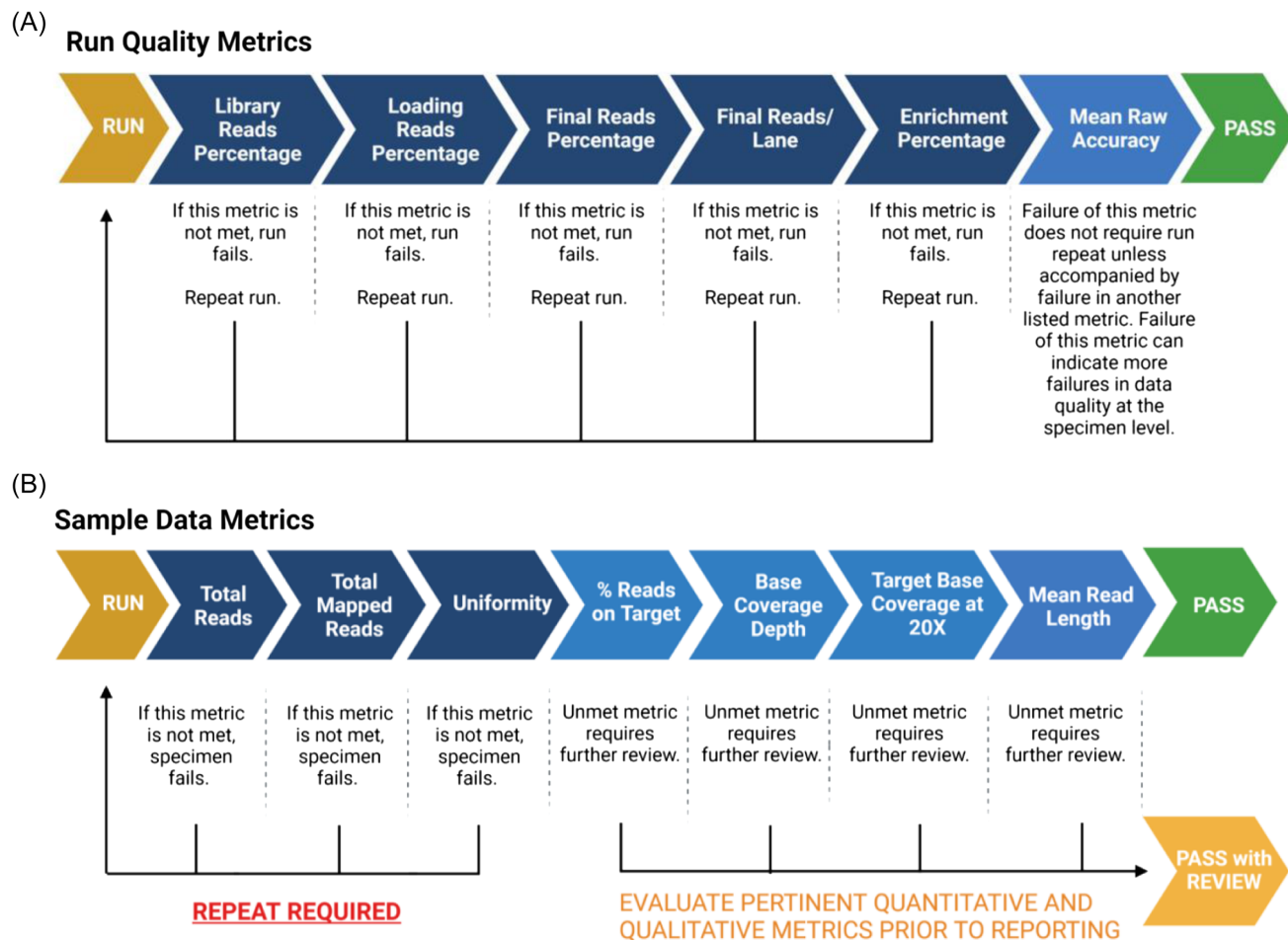


Figure 2. Quality metrics application workflow overall metric application of both run and sample data metrics were used for every Genexus run and specimen. Run metrics were processed as described (A). Failure in all but 1 of the designated run metrics ended the analysis and requires repeat sequencing to obtain reportable data. Individual sequence data were evaluated (B) thoroughly before final variant determination. Failure in any listed metric either ended analysis or required further assessment prior to reporting.

workflow may incorporate automated parsing of base and mapping quality scores to enhance the resolution and reliability of SARS-CoV-2 variant detection.

RESULTS

Quantification of Samples

Accurate quantification of viral copy number or RNA concentration was essential to ensure appropriate assay selection and successful sequencing outcomes. We compared 2 quantification methods recommended by Thermo Fisher: the TaqPath COVID-19 Combo Kit and the Qubit Fluorometer. Despite its use as an RNA quantification tool, multiple extractions analyzed on the Qubit Fluorometer failed to detect sufficient nucleic acid, often resulting in “Too Low” readings, even for samples that yielded successful sequencing results on the Genexus Integrated Sequencer. Additionally, when comparing the Ct values from TaqPath quantification to Qubit readings, there was no observable correlation, further bringing into

question the reliability of the Qubit Fluorometer for this application. After extensive testing, we determined that all properly processed samples with an N gene Ct value of at least 30 consistently produced successful sequence annotations on the Ion Torrent Genexus platform. This observation confirmed the TaqPath COVID-19 Combo Kit as our preferred quantification method because of its reliability and alignment with sequencing success.

Controls

In total, 431 specimens were sequenced across 32 runs, including 89 control samples: NTCs ($n = 14$) and positive commercial reference material (CRM) controls ($n = 75$). Positive CRM controls were diluted to concentrations both above and below the instrument’s minimum detectable range, specifically at 250 \times , 10 \times , 3 \times , 1 \times , and 0.3 \times of the LOD. Positive control material was included to confirm the method’s ability to accurately sequence existing SARS-CoV-2 variants. NTC samples consisted of

uninoculated VTM ($n = 5$) and CRM controls for non-COVID-19 coronaviruses (strains 229E, HKU1, NL63, and OC43) ($n = 9$), each processed similarly to patient samples to ensure uniformity in testing. Negative material was included to provide periodic control for contamination.

To define minimum QC thresholds, we organized and evaluated positive control metrics (Figure 3). The following QC metrics were initially informed by manufacturer documentation but were refined through empirical analysis of our control data set. “Passing” sequences were required to meet the GenBank minimum base requirement (29 400 bases) and yield accurate lineage assignments (Supplementary Table 1). The most variable metrics among controls were total reads and number of mapped reads, leading us to establish more inclusive thresholds of 100 000 bases and 50 000 bases, respectively, for these categories. This adjustment enabled 59 of 89 controls to meet the “Pass” criteria. Notably, 3 Twist controls failed the base pair requirement, as these controls are inherently incomplete sequences. Among controls below the LOD, 5 failed to meet the base pair threshold and were assigned incorrect lineages, while 1 control achieved correct lineage assignment despite falling below the base pair threshold. Additionally, 3 negative control sequences unexpectedly passed the QC metrics, with 2 producing false-positive results and the third failing to generate a FASTA file for alignment. For sequences that “Passed” QC yet were incorrectly aligned, target coverage at 20× was below 90%, whereas sequences with correct lineage assignments had target coverage at or above 95%.

In our 32 sequencing runs, 14 patient specimens previously confirmed as SARS-CoV-2 negative were used as negative controls, each with N gene Ct values of “Undetected” or exceeding the manufacturer’s recommended threshold of 37. Of these NTCs, 11 failed QC metrics as expected, while 3 unexpectedly passed. The failing samples included 4 uninoculated VTM samples and 7 non-COVID-19 coronavirus controls. Among the VTM samples, 2 did not produce FASTA files, while 3 erroneously received lineage assignments in PANGOLIN, likely due to false-positive calls, highlighting the need for robust QC filtering before further data analysis.²⁵⁻³⁰ For the non-COVID-19 coronavirus controls, 1 sample produced a FASTA file and was assigned a lineage, while 5 others failed QC metrics in PANGOLIN. Of the 2 NTC samples that passed QC and were assigned lineages, one displayed a low viral copy number during quantification, and the other’s lineage assignment was likely due to laboratory scientist error, as later confirmed by reanalysis on the SARS-CoV-2 Insight High Titer Assay. These findings emphasize the importance of careful handling and validation when interpreting NTC results, as minor deviations can lead to erroneous lineage assignments. In addition, materials without any potential viral material, such as uninoculated VTM, should be used as NTCs, to avoid false-positive results or low-titer calls that could confound results and delay variant tracking during troubleshooting.

Patient Specimens

In addition to controls, 339 patient specimens and 3 proficiency testing samples were analyzed using the Genexus platform. Of these, 304 samples met both run and sample QC criteria, which included metrics for the SARS-CoV-2 Research Panel, the SARS-CoV-2 Insight Panel High Titer assay, and the SARS-CoV-2 Insight Panel Low Titer assay. N gene Ct values for these samples ranged between 12 and 32 for the Research Panel, 11 and 29 for the high-titer assay, and 13 and 33 for the low-titer assay. Among the 38 specimens that failed QC, 17 exhibited failures at the run level, indicating overall poor performance that precluded further data analysis. The remaining 21 failed at the sample level, with N gene Ct values from 15 to “Undetected.” Most failures in this group had Ct values above 28, the manufacturer’s upper limit for the SARS-CoV-2 Insight High Titer Assay and the sequencing threshold recommended by the Centers for Disease Control and Prevention (CDC).³¹ Reasons for failure included low eluate volume in 8 specimens as well as issues related to sample degradation from repeated freeze-thaw cycles or multiple extraction attempts ($n = 7$) and potential loading or handling errors ($n = 6$). These failed specimens demonstrated deficiencies across metrics such as total reads, mapped reads, uniformity, target base coverage at 20×, or a combination of these parameters, underscoring the need for rigorous adherence to quality metrics in sample preparation.

During the testing period, the Delta variant emerged as the predominant strain, representing approximately two-thirds of all cases analyzed (Figure 4A). Initially, the Alpha variant was most prevalent, but from Summer through Autumn 2021, Delta became the dominant variant in our population sample (Figure 4B). Our laboratory’s sequencing surveillance allowed for early detection of the Omicron variant in late December 2021. Because of Omicron’s increased transmissibility, this variant rapidly replaced Delta as the primary strain circulating in the population.

NGS Metrics and Analysis

To assess the quality of sequencing runs and individual samples, metrics illustrated in Figure 2 were applied. Sample metrics were only considered if the corresponding run metrics met QC thresholds. Although sample quality could be deemed acceptable with lower values for certain categories (as detailed in Table 1), qualitative metrics were necessary for specimens not meeting all “Pass” thresholds. Specimens that satisfied these QC criteria showed minimal gaps and met the Global Initiative on Sharing All Influenza Data (GISAID) and GenBank 90% genome coverage requirement. All patient sequences that passed QC had a minimum sequence length of 29 400, with an average length of 29 784, thus fulfilling GenBank’s sequence length requirements.³²

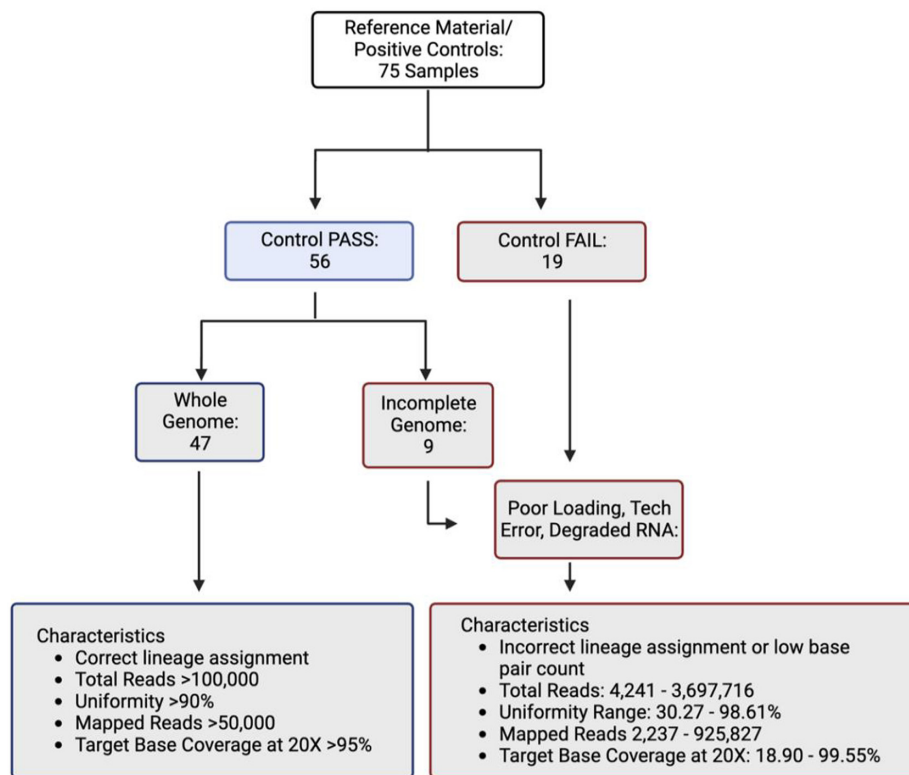


Figure 3. Commercial reference material control failures and implications. At present, 75 positive controls at various concentrations above and below the limit of detection have been sequenced. Using the metrics applied in Figure 2, 27 control sequences “Pass,” yielding a high confidence in the accuracy and future application of these sequences. These specimens were loaded at concentrations above $3\times$ the assay limit of detection and exhibited relatively high total reads, uniformity, and reads on target. Those that fail display larger gaps in the consensus sequence, weakening the overall quality of the sequences despite some of their accurate lineage classifications. Controls that failed had a wide range of metric readings, with the lower limits of total reads, uniformity, and reads on target reaching 4241, 30.27%, and 7.55%, respectively.

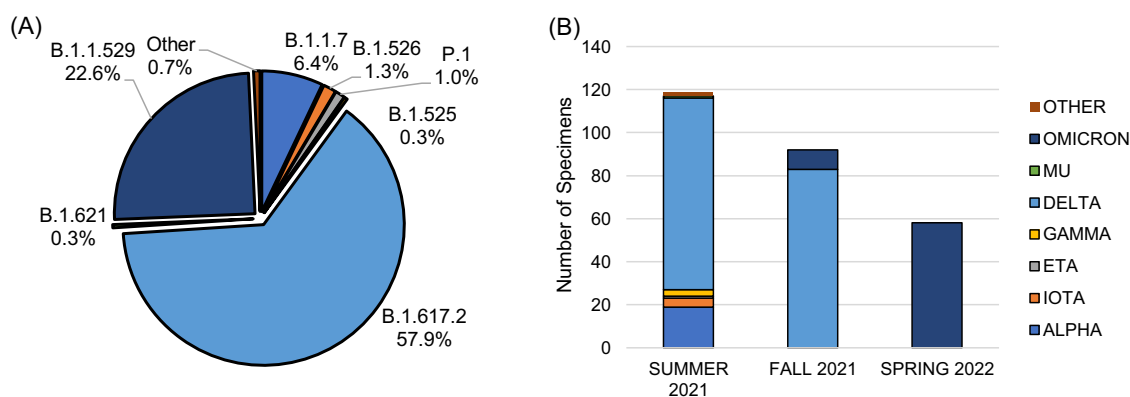


Figure 4. Prevalence of SARS-CoV-2 variants from May 2021 to April 2022. During the active sequencing timeframe, B.1.617.2 (Delta) was the dominant variant subjected to sequencing workflow, with B.1.1.529 (Omicron) and B.1.1.7 (Alpha) posing a strong presence (A). Alpha was the dominant variant from May to June of 2021, with Delta quickly taking this position from July to December of 2021. Omicron made its first appearance on campus in December 2021 and remains the most prevalent variant on campus as of April 2022 (B).

DISCUSSION

As new SARS-CoV-2 variants emerge, it is critical to maintain accurate lineage assignments to support variant tracking and clinical reporting. Reliable assignments require robust QC metrics that ensure sequencing data are accurately interpreted, producing high-quality sequences suitable for annotation and other downstream applications. Although several sequencing platforms are available for SARS-CoV-2, standardized quality metrics are yet to be widely adopted for the Ion Torrent Genexus Integrated Sequencer. In this study, we proposed a set of QC metrics that, when applied, accurately resolved 100% of third-party positive control specimens meeting these metrics, thereby minimizing gaps in the consensus sequence, reducing errors, and mitigating issues associated with low-quality sequences. Through consultations with manufacturer bioanalysts and analysis of SARS-CoV-2 NGS data from hundreds of specimens, we identified clinical metrics directly associated with sequence quality. These metrics, as detailed in Table 1 and illustrated in Figure 2, offer a structured QC workflow that enables clinical staff to efficiently identify and address potential quality issues.

The initial step in sequencing involves RNA extraction from controls or patient specimens. Thermo Fisher recommends using the TaqPath COVID-19 Combo Kit for semi-quantitative viral copy assessment. Although the Qubit Fluorometer was also suggested as an alternative quantification tool, we observed inconsistent RNA concentration measurements and no correlation with Ct values. Similar findings have been reported by other studies, indicating potential limitations of the Qubit Fluorometer for this application.³³ While the Qubit method offers faster quantification, our data indicated it was unreliable for the sequencing workflow. Thus, we adopted Ct values from the TaqPath assay to determine the appropriate assay selection (low titer or high titer) and any necessary dilutions. Although the assay package insert and CDC guidelines recommend a Ct threshold of 28 for sequencing, we observed successful sequencing at Ct values up to 31 when using the SARS-CoV-2 Insight Panel. Additional studies and specimens would be required to further explore the possibility of consistently successful sequencing of SARS-CoV-2 at high Ct values.

Approximately 17.5% of the specimens analyzed in this study ($n = 75$) were commercially available positive controls representing 5 viral variants (Alpha, Beta, Delta, Gamma, and Omicron). Of these, 47 met GenBank's genome requirements and received correct lineage assignments, with each sample meeting thresholds of at least 100 000 total reads, at least 50 000 mapped reads, at least 95% target base coverage at 20 \times , and at least 90% uniformity. Specimens near the lower threshold for total and mapped reads required high uniformity and target base coverage to meet the minimum base pair requirement and achieve accurate lineage assignment. Of the positive controls that failed the QC metrics, 27 did not

meet GenBank's minimum base requirement, and the remaining control fell below the total reads threshold, demonstrating that the use of these metrics maintains a standard of adequate SARS-CoV-2 genome sequencing, not partial or incomplete sequencing. This finding underscores the importance of including base pair count as part of the QC criteria to ensure analysis on complete genomes. Adherence to these metrics provides confidence in the integrity and completeness of the generated sequences.

Among the negative controls, 86% performed as expected, while 2 out of 14 unexpectedly passed QC metrics and received lineage assignments with full genome base pair counts. Four blank specimens failed individual QC metrics; however, 3 of the 5 blank samples received an erroneous lineage assignment in PANGOLIN. This phenomenon, known as false lineage assignment or positive calling, is documented in the Ion AmpliSeq SARS-CoV-2 Insight Panel User Guide and has been previously reported in the literature on Ion Torrent platforms for whole-exome and targeted sequencing.²⁸⁻³⁰ The false assignment of lineages in specimens that consistently tested negative for SARS-CoV-2 through diagnostic assays and RNA quantification highlights the necessity of rigorous QC filtering before using data in downstream applications. Of the negative controls, 9 were positive for non-COVID-19 coronaviruses; 2 of these controls passed the QC metrics and were incorrectly assigned a lineage. Among the 3 negative controls that "Passed" metrics, 1 failed to produce a FASTA file, while the other 2 were assigned incorrect lineages and were initially processed on the SARS-CoV-2 Insight Low Titer assay. Upon retesting with the SARS-CoV-2 Insight High Titer assay, 1 of these samples appropriately failed according to the proposed QC metrics. These findings underscore 3 critical considerations: contamination risk, importance of assay selection, and the need for careful screening of presumed negative specimens. In 1 instance, contamination likely resulted from laboratory scientist error, as the incorrectly assigned lineage matched that of a previously positive sample on the same plate. Further testing confirmed that RNA contamination was absent in the extraction plate, as retesting yielded appropriate failure. In another case, the second non-COVID-19 control produced late Ct values during qPCR, indicating a viral copy number below the LOD of both the TaqPath COVID-19 Combo Kit and BioFire Respiratory 2.1 (RP2.1) Panel. It is plausible that this low viral material was successfully sequenced on the low-titer assay because of the additional PCR cycles during library preparation. These sequences contained large gaps in the consensus sequence when aligned to the SARS-CoV-2 reference genome, further affirming the need for careful evaluation of low-quality samples. The complications detailed here further support the use of uninoculated VTM as an NTC instead of previously negative patient specimens, as low-titer specimens, host RNA, or potentially closely related pathogens may be amplified by sequencing, thus delaying variant reporting during troubleshooting.

Specimens were loaded and sequences were analyzed using the manufacturer's recommendations and software, respectively, as this was an assessment of the method as a measure of efficiency in SARS-CoV-2 sequencing. Additional analysis of output data could prove useful but was not performed at this time.

While this study proposes a robust framework for evaluating SARS-CoV-2 sequencing data on the Ion Torrent Genexus platform, several limitations should be acknowledged. Although the system calculates base and mapping quality scores using modified Phred scales, their limited accessibility within the SARS-CoV-2 Insight Panel workflow prevented direct incorporation into our QC thresholds. Instead, we relied on mapped reads, genome coverage, and uniformity metrics to assess data quality. For samples requiring further scrutiny, consensus alignments were visually inspected for gaps or inconsistencies. Future refinements to the workflow may incorporate automated parsing of base and mapping quality metrics to enhance downstream variant call confidence. Another important consideration relates to the intrinsic limitations of the Ion Torrent sequencing platform. Homopolymeric stretches and high GC-content regions are prone to sequencing artifacts such as indel errors and coverage dropout. Although our analysis evaluated genome completeness and depth, we did not perform regional assessments of error distribution across the viral genome. Consequently, variants occurring in these challenging sequence contexts may be underreported or misclassified. Additionally, off-target amplification of host RNA, particularly in low-titer or degraded samples, may introduce noise or lead to false-positive variant calls. While not filtered in this study, future workflows could benefit from incorporating host depletion strategies or bioinformatic screening to remove nonviral reads prior to analysis. Finally, this study did not perform explicit trimming of primer sequences from the sequencing reads, which may reduce sensitivity for detecting variants near primer binding sites. Future pipelines should assess the impact of primer trimming on variant detection, particularly as new SARS-CoV-2 lineages emerge with mutations in these regions.

During our study, Thermo Fisher introduced the SARS-CoV-2 Insight Assay to replace the original SARS-CoV-2 Research Panel. The QC metrics established here were applicable to both assays. Several specimens initially tested with the Research Panel were subsequently retested with the Insight assay, consistently yielding identical lineage assignments across assays. Notably, the AccuPlex Seracare Control reliably generated correct lineage assignments when sequence data passed the proposed QC metrics. Given that these assays produce near-complete sequences in successful runs, applying these metrics should not exclude novel variants from annotation, even with the rapid mutation rate seen in Omicron.^{34,35} Metric failure is generally indicative of overall run failure or poor nucleic acid quality, which can lead to significant gaps in the sequence. Further research is

needed to validate these metrics for other NGS platforms, viral targets, or primer panels, as their applicability may vary across different instruments and sample types. Any laboratory intending to implement these metrics should conduct thorough validation to address lab-specific requirements. Additionally, any panel modifications or software updates that impact analysis necessitate performance verification to ensure continued compliance with CLIA guidelines for laboratory-developed tests.

In conclusion, the QC metrics proposed in this study enable accurate lineage assignments with minimal gaps in the consensus sequence. Of the positive control specimens that failed, 27 did not meet the whole genome requirement, with 1 sample falling below the total reads threshold. The failed positive controls exhibited sequence gaps that could adversely affect lineage assignment and data application. Consistently, specimens with matching variants generated repeatable lineage assignments on runs with passing QC metrics. Collectively, these findings support the use of the proposed QC metrics as a standardized approach for SARS-CoV-2 sequencing on the Genexus Integrated Sequencer, offering a potential model for uniform QC across laboratories.

REFERENCES

- Han P, Li L, Liu S, et al. Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. *Cell*. 2022;185(4):630–640.e610. doi: [10.1016/j.cell.2022.01.001](https://doi.org/10.1016/j.cell.2022.01.001)
- Harvey WT, Carabelli AM, Jackson B, et al; COVID-19 Genomics UK (COG-UK) Consortium. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*. 2021;19(7):409–424. doi: [10.1038/s41579-021-00573-0](https://doi.org/10.1038/s41579-021-00573-0)
- Tao K, Tzou PL, Nouhin J, et al. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat Rev Genet*. 2021;22(12):757–773. doi: [10.1038/s41576-021-00408-x](https://doi.org/10.1038/s41576-021-00408-x)
- Avendano C, Lilienfeld A, Rulli L, et al. SARS-CoV-2 variant tracking and mitigation during in-person learning at a Midwestern university in the 2020–2021 school year. *JAMA Netw Open*. 2022;5(2):e2146805. doi: [10.1001/jamanetworkopen.2021.46805](https://doi.org/10.1001/jamanetworkopen.2021.46805)
- Pollock BH, Kilpatrick AM, Eisenman DP, et al. Safe reopening of college campuses during COVID-19: the University of California experience in Fall 2020. *PLoS One*. 2021;16(11):e0258738. doi: [10.1371/journal.pone.0258738](https://doi.org/10.1371/journal.pone.0258738)
- Alessandrini F, Caucci S, Onofri V, et al. Evaluation of the Ion AmpliSeq SARS-CoV-2 Research Panel by massive parallel sequencing. *Genes (Basel)*. 2020;11(8):929. doi: [10.3390/genes11080929](https://doi.org/10.3390/genes11080929)
- Liu T, Chen Z, Chen W, et al. A benchmarking study of SARS-CoV-2 whole-genome sequencing protocols using COVID-19 patient samples. *iScience*. 2021;24(8):102892. doi: [10.1016/j.isci.2021.102892](https://doi.org/10.1016/j.isci.2021.102892)
- Kubik S, Marques AC, Xing X, et al. Recommendations for accurate genotyping of SARS-CoV-2 using amplicon-based sequencing of clinical samples. *Clin Microbiol Infect*. 2021;27(7):1036.e1–1036.e8. doi: [10.1016/j.cmi.2021.03.029](https://doi.org/10.1016/j.cmi.2021.03.029)
- Lim HG, Hsiao SH, Fann YC, Lee YG. Robust mutation profiling of SARS-CoV-2 variants from multiple raw illumina sequencing data with Cloud Workflow. *Genes (Basel)*. 2022;13(4):686. doi: [10.3390/genes13040686](https://doi.org/10.3390/genes13040686)

10. Bruzek S, Vestal G, Lasher A, Lima A, Silbert S. Bacterial whole genome sequencing on the Illumina iSeq 100 for clinical and public health laboratories. *J Mol Diagn.* 2020;22(12):1419–1429. doi: [10.1016/j.jmoldx.2020.09.003](https://doi.org/10.1016/j.jmoldx.2020.09.003)
11. Huang B, Jennison A, Whitley D, et al. Illumina sequencing of clinical samples for virus detection in a public health laboratory. *Sci Rep.* 2019;9(1):5409. doi: [10.1038/s41598-019-41830-w](https://doi.org/10.1038/s41598-019-41830-w)
12. Fissel JA, Mestas J, Chen PY, et al. Implementation of a streamlined SARS-CoV-2 whole-genome sequencing assay for expeditious surveillance during the emergence of the Omicron variant. *J Clin Microbiol.* 2022;60(4):e0256921. doi: [10.1128/jcm.02569-21](https://doi.org/10.1128/jcm.02569-21)
13. Kulkarni S, Pfeifer J. *Clinical Genomics.* Elsevier/Academic Press; 2015.
14. Pillay S, Giandhari J, Tegally H, et al. Whole genome sequencing of SARS-CoV-2: adapting illumina protocols for quick and accurate outbreak investigation during a pandemic. *Genes (Basel).* 2020;11(8):949. doi: [10.3390/genes11080949](https://doi.org/10.3390/genes11080949)
15. Plitnick J, Griesemer S, Lasek-Nesselquist E, et al. Whole-genome sequencing of SARS-CoV-2: assessment of the Ion Torrent AmpliSeq Panel and comparison with the Illumina MiSeq ARTIC protocol. *J Clin Microbiol.* 2021;59(12):e0064921. doi: [10.1128/JCM.00649-21](https://doi.org/10.1128/JCM.00649-21)
16. Sundin T, Aunchman M, Grissom L, et al. Comparison of next-generation sequencing assays for clinical use in solid tumor malignancies. *J Mol Diagn.* 2019;1183.
17. Hofman P, Bordone O, Chamorey E, et al. Setting-up a rapid SARS-CoV-2 genome assessment by next-generation sequencing in an academic hospital center (LPCE, Louis Pasteur Hospital, Nice, France). *Front Med (Lausanne).* 2021;8:730577. doi: [10.3389/fmed.2021.730577](https://doi.org/10.3389/fmed.2021.730577)
18. Hill S, Perkins M, von Eije KJ, et al. *Genomic Sequencing of SARS-CoV-2: A Guide to Implementation for Maximum Impact on Public Health.* World Health Organization; 2021.
19. Firew T, Sano ED, Lee JW, et al. Protecting the front line: a cross-sectional survey analysis of the occupational factors contributing to healthcare workers' infection and psychological distress during the COVID-19 pandemic in the USA. *BMJ Open.* 2020;10(10):e042752. doi: [10.1136/bmjopen-2020-042752](https://doi.org/10.1136/bmjopen-2020-042752)
20. High hopes for schools as lab positions go unfilled. *CAP TODAY.* May 2022. Accessed June 12, 2025. <https://www.captodayonline.com/high-hopes-for-schools-as-lab-positions-go-unfilled/>.
21. Zhang SX, Chen J, Afshar Jahanshahi A, et al. Succumbing to the COVID-19 pandemic-healthcare workers not satisfied and intend to leave their jobs. *Int J Ment Health Addict.* 2022;20(2):956–965. doi: [10.1007/s11469-020-00418-6](https://doi.org/10.1007/s11469-020-00418-6)
22. Lahens NF, Ricciotti E, Smirnova O, et al. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genomics.* 2017;18(1):602. doi: [10.1186/s12864-017-4011-0](https://doi.org/10.1186/s12864-017-4011-0)
23. Rümke LW, Groenveld FC, van Os YMG, et al. In-depth characterization of vaccine breakthrough infections with SARS-CoV-2 among health care workers in a Dutch academic medical center. *Open Forum Infect Dis.* 2021;9(1):ofab553. doi: [10.1093/ofid/ofab553](https://doi.org/10.1093/ofid/ofab553)
24. Jacot D, Pillonel T, Greub G, Bertelli C. Assessment of SARS-CoV-2 genome sequencing: quality criteria and low-frequency variants. *J Clin Microbiol.* 2021;59(10):e0094421. doi: [10.1128/JCM.00944-21](https://doi.org/10.1128/JCM.00944-21)
25. Ion AmpliSeq™ SARS-CoV-2 Insight Research Assay – GX: user guide. Thermo Fisher. 2025. Accessed June 12, 2025. https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FMSG%2Fmanuals%2FMAN0024933_1AInsight-CoV2forGX_UG.pdf.
26. Genexus™ Integrated Sequencer: user guide. Thermo Fisher. 2025. Accessed June 12, 2025. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017910_Genexus-IntegratedSequencer_UG.pdf.
27. Ion AmpliSeq™ SARS-CoV-2 Research Panel. Thermo Fisher. 2020. Accessed June 12, 2025. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0019278_Ion_AmpliSeq_SARS-CoV-2_Research_Panel_Genexus_QR.pdf.
28. Gouvêa Moreira TC, Da Silva Spinola P, Campos Rezende M, et al. Correlation between the number of false positive variants and the quality of results using Ion Torrent PGM™ sequencing to screen BRCA genes. *Biomedica.* 2021;41(4):773–786. doi: [10.7705/biomedica.5663](https://doi.org/10.7705/biomedica.5663)
29. Gampawar P, Saba Y, Werner U, Schmidt R, et al. Evaluation of the performance of AmpliSeq and SureSelect exome sequencing libraries for Ion Proton. *Front Genet.* 2019;10:856. doi: [10.3389/fgene.2019.00856](https://doi.org/10.3389/fgene.2019.00856)
30. Damiani E, Borsani G, Giacomuzzi E. Amplicon-based semiconductor sequencing of human exomes: performance evaluation and optimization strategies. *Hum Genet.* 2016;135(5):499–511. doi: [10.1007/s00439-016-1656-8](https://doi.org/10.1007/s00439-016-1656-8)
31. Wentworth DE. *Updated CDC Guidance for Specimen Submission for Surveillance of SARS-CoV-2 (NS3).* Centers for Disease Control and Prevention; 2022.
32. Recommendations for SARS-CoV-2 sequence data quality & reporting. APHL. March 1, 2021. Accessed June 12, 2025. <https://www.aphl.org/programs/preparedness/Crisis-Management/Documents/APHL-SARS-CoV-2-Sequencing.pdf>.
33. Rachiglio AM, De Sabato L, Roma C, et al. SARS-CoV-2 complete genome sequencing from the Italian Campania region using a highly automated next generation sequencing system. *J Transl Med.* 2021;19(1):246. doi: [10.1186/s12967-021-02912-4](https://doi.org/10.1186/s12967-021-02912-4)
34. Chatterjee S, Kim C-M, Lee YM, et al. Whole-genome analysis and mutation pattern of SARS-CoV-2 during first and second wave outbreak in Gwangju, Republic of Korea. *Sci Rep.* 2022;12(1):11354. doi: [10.1038/s41598-022-14989-y](https://doi.org/10.1038/s41598-022-14989-y)
35. Wang S, Xu X, Wei C, et al. Molecular evolutionary characteristics of SARS-CoV-2 emerging in the United States. *J Med Virol.* 2022;94(1):310–317. doi: [10.1002/jmv.27331](https://doi.org/10.1002/jmv.27331)