Application Note

Rapid detection of SARS-CoV-2 variants by automated library preparation and next-generation sequencing

A demonstrated workflow for QIAseq® DIRECT Enhanced SARS-CoV-2 library preparation on the Sciclone® NGSx iQ™ workstation

Introduction

The rapid spread of SARS-CoV-2 has led to the emergence of a series of novel variants, such as the Alpha (B.1.1.7), Beta (B.1.351), Gamma (P1, Brazil), Delta (B.1.617.2) and Omicron (B.1.1.529). A comprehensive understanding of mutations in new variants, particularly those that can heighten transmissibility and severity of illness, is critical in the fight against the virus.

Next-generation sequencing (NGS) is the method of choice for detecting emerging variants and estimating their epidemiological prevalence. The SARS-CoV-2 genome is relatively small (30 kb RNA) and can be detected in various samples, including nasopharyngeal swabs, buccal swabs and wastewater. A typical sequencing workflow starts with RNA isolation, followed by cDNA synthesis, target enrichment, fragmentation of target-enriched libraries, adapter ligation, library amplification and sequencing. Since the pandemic's beginning, most primer designs have been based on primer sequences from the ARTIC network (1), amplifying 400 bp amplicons into two PCR pools covering the entire SARS-CoV-2 genome. However, this approach is long and laborious, requiring fragmentation of the target-enriched libraries. In addition, there is an increased risk of dropouts and missed variants if the primers do not anneal to the 30 kb template of the virus.

The QIAseq DIRECT SARS-CoV-2 protocol introduces a streamlined, four-hour enrichment and library preparation workflow. It removes the need for fragmentation and quantification/normalization before library amplification and indexing. This reduces library preparation time and allows more samples to be multiplexed on one sequencing run. Moreover, the workflow maintains robustness and uniformity of coverage and provides better variant discovery than the ARTIC network-based protocols.

Even with a streamlined workflow, manual library prep can be tedious and inefficient, requiring long hands-on and incubation times. In addition, steps like magnetic bead washing to remove PCR remains can cause low overall library yields. Automating the workflow can reduce operational costs, error rates and sample variability and allow labs to reliably process small or large batches of samples simultaneously. ▷

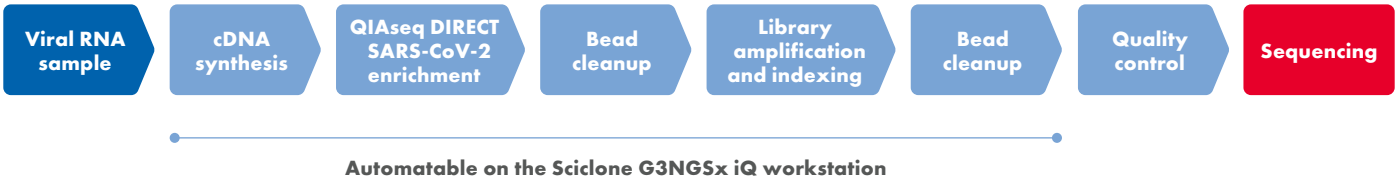

Figure 1. Sample to sequencing workflow showing the automatable library preparation steps.

In this Application Note, we introduce a plug-and-play library preparation workflow (Figure 1) for detecting SARS-CoV-2 variants using the QIAseq DIRECT enhanced protocol automated on PerkinElmer's Sciclone G3NGSx iQ workstation for sequencing on Illumina's NextSeq® 2000. The workstation's on-deck thermocycler and small-volume liquid handling capabilities allow users to perform the whole process hands-free in a six-hour workflow.

Materials and methods

Viral RNA extraction: As per the manufacturer's instructions, 400 µl of each sample was eluted to 50 µl on the KingFisher™ Flex Purification System (ThermoFisher, Waltham, MA).

NGS library preparation: For manual library preparation, 12 samples were processed, and for automated library preparation, 80 samples were processed. cDNA was synthesized from 5 µl of extracted RNA using the QIAseq DIRECT Enhanced SARS-CoV-2 Library Kit. A total of 16 µl from the cDNA reaction was prepared with two amplicon pools. The pools were combined, and 24 µl of the target enrichment reaction was amplified and sample-indexed. For the automated workflow, the enriched sequencing libraries were generated on the Sciclone G3 NGSx iQ in a high-throughput method. The user interface of the workbench prompts users to select the required number of columns to process from 8–96 samples at a time and review the correct deck set up for the workflow (Figure 2). Master mixes and reagents were prepared as shown in Figure 3. Table 1 compares the time taken to complete the two workflows.

Table 1. A comparison of manual vs. automated library prep workflow

Step	Manual workflow	Automated workflow
cDNA synthesis	75 minutes	75 minutes
QIAseq DIRECT SARS-CoV-2 enrichment	145 minutes	130 minutes
Bead cleanup	45 minutes	40 minutes
Library amplification and indexing	45 minutes	45 minutes
Bead cleanup	45 minutes	40 minutes
Total time	6 hours	5.5 hours

Quality control and next-generation sequencing: The quality and quantity of libraries were assessed using a subset of samples. 5 µl of each library was pooled and sequenced on an Illumina NextSeq 2000 instrument with high-output flow cells, generating 150 bp paired-end reads. The target for sequencing was roughly 125,000 read clusters for each sample.

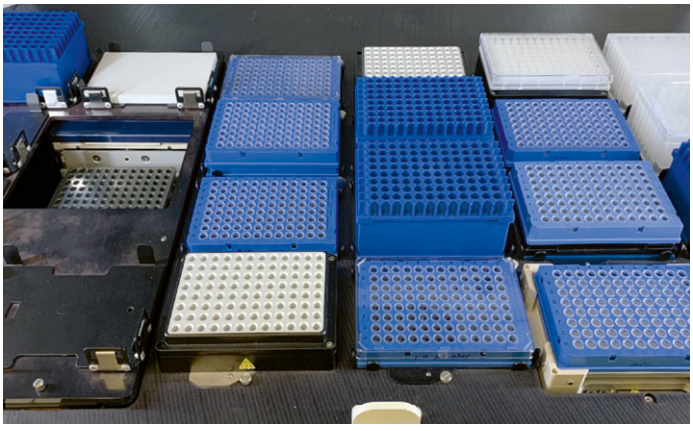

Figure 2. Example of the deck layout for the QIAseq DIRECT Enhanced protocol setup during cDNA synthesis – the first step of the workflow on the Sciclone G3 NGSx iQ workstation.

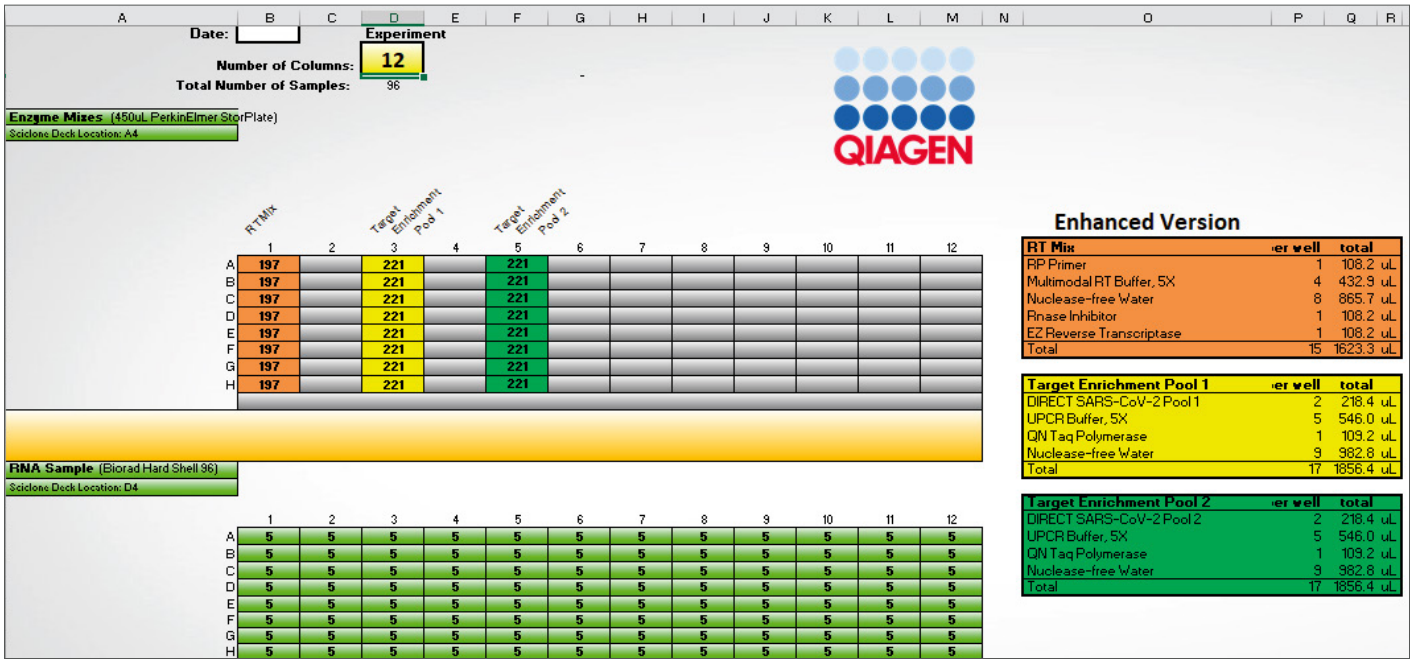

Figure 3. An example of the Excel® workbook describing the recipes and volumes required for the different master mixes. The number of columns can be adjusted to support a variety of sample quantities throughout the automated workflow.

Results

Here we show successful and reliable automation of QIAseq DIRECT Enhanced SARS-CoV-2 library preparation on the PerkinElmer Sciclone NGSx iQ workstation.

The library yield was higher, and the overall median library concentration and region molarity were almost double for the automated workflow compared to the manual workflow (Figure 4). A comparison of final libraries shows strong library peaks at about 399 bp for the automated workflow compared to varied library peaks at 400 bp and primer dimer peaks at ~180 bp for the manual workflow (Figure 5). These results show that, while the two methods are comparable, library generation using the automated workflow is more efficient.

The automated workflow allows samples with a wide range of C_T values to be successfully multiplexed in a simple run. Coverage greater than 20X, a common requirement for SARS-CoV-2 sequencing, was observed for all samples irrespective of their initial C_T values, which ranged from ~17–35 (Figure 6). This eliminates the need for batching samples based on C_T value to obtain sequencing data with >80% of total reads mapping to the SARS-CoV-2 genome (Figure 4 A). Additionally, achieving high coverage in the spike protein region of the SARS-CoV-2 genome is challenging. This is not true in our case – except for one, all samples achieved high coverage (more than 20% of the reads gaining 100X coverage), which far exceeds the 20X coverage standard (Figure 7). The median number of reads for the run was 321,810 (31,176, 2,197,842).

Moreover, more reads per sample mean better overall coverage of the genome. Figure 8 shows this positive correlation between coverage and reads from the automated library prep workflow. Despite some samples getting significantly less than the recommended 125,000 reads per sample, the median coverage for all samples on the run was >1000. This demonstrates that highly sensitive and reproducible data can be generated even with a significantly lower number of reads and varying viral titer. ▷

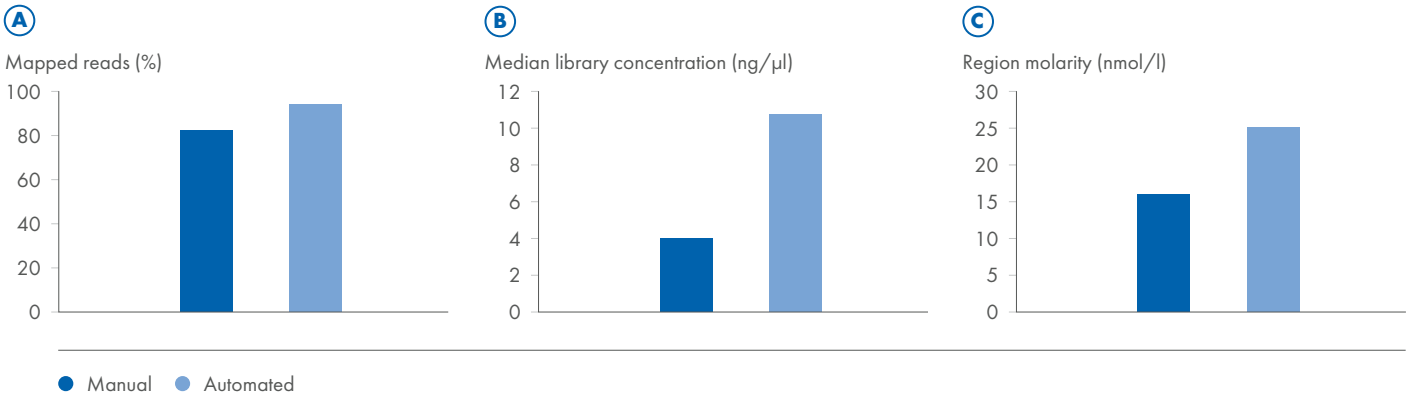

Figure 4. Comparison of sequencing results – manual vs automated. A Percentage of reads mapped to the SAR-CoV-2 genome. **B** Median library concentration (ng/µl). **C** Region molarity (nmol/l) of final libraries. The expected library size was approximately 400 bp, and the region molarity was measured between 270–750 bp.

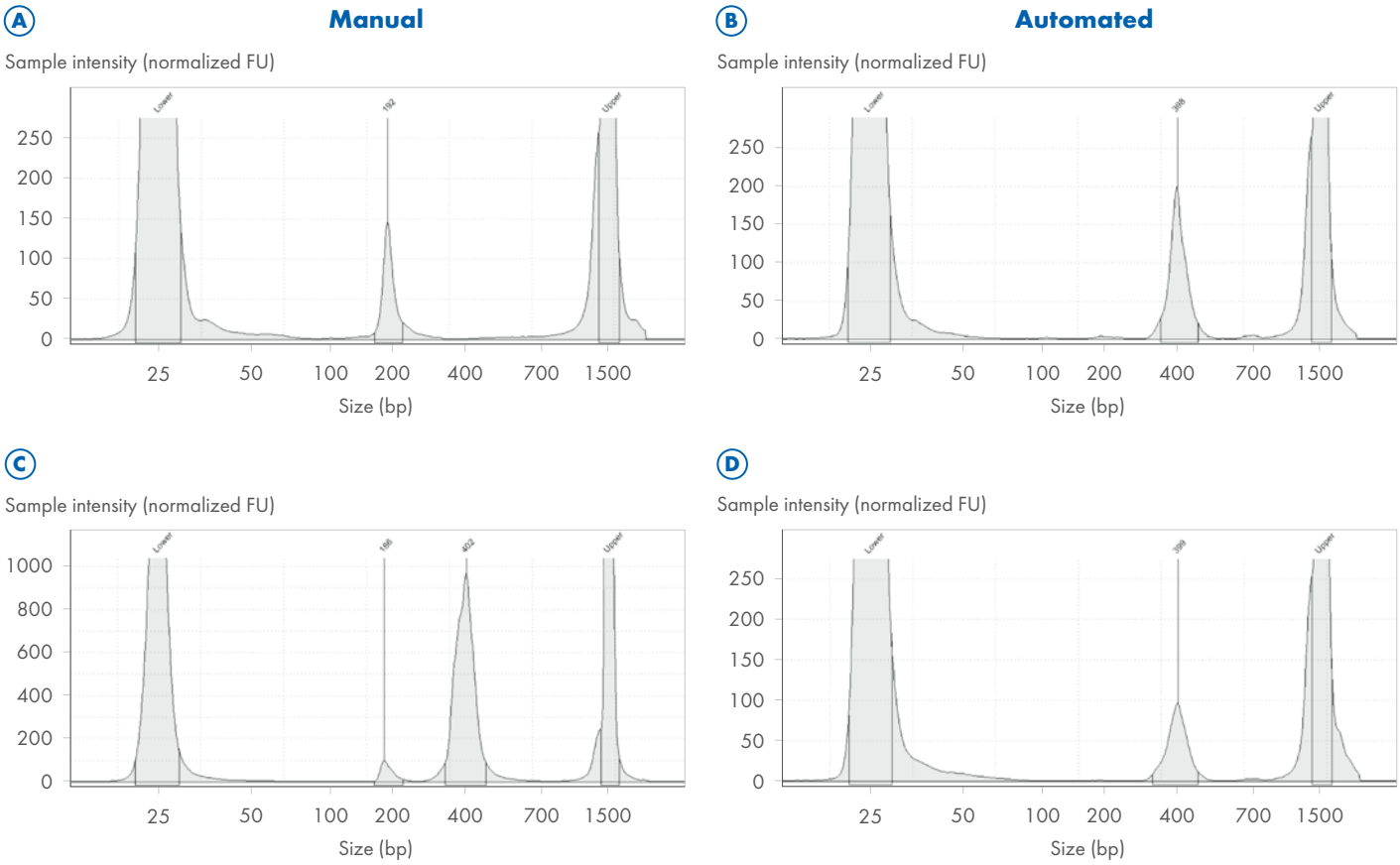

Figure 5. Comparison of library traces – control RNA-manual vs. control RNA-automated.

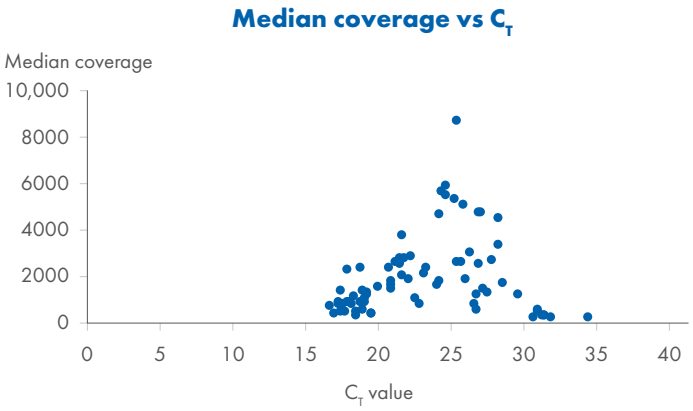

Figure 6. Sequence data mapping the coverage vs. the C_T of the input samples before cDNA synthesis. This shows coverage greater than 20X for all samples despite varying initial C_T values ranging from ~17–35.

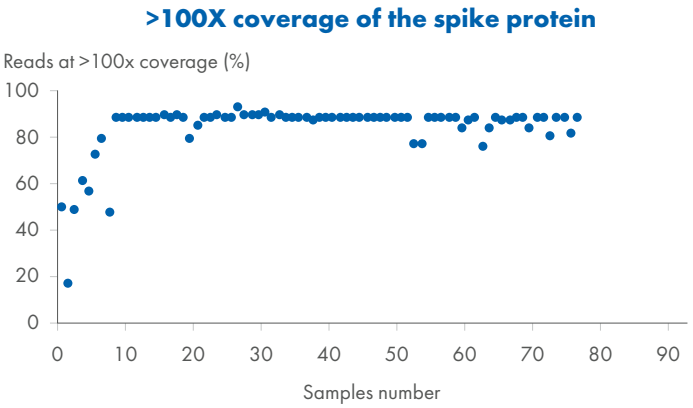

Figure 7. Sequencing data measuring the percentage of the reads mapping to the spike protein with a coverage greater than 100X. The general standard for acceptable coverage of sequencing of SARS-CoV-2 is measured by the % of reads having >20X coverage. This depiction shows coverage far exceeding that standard.

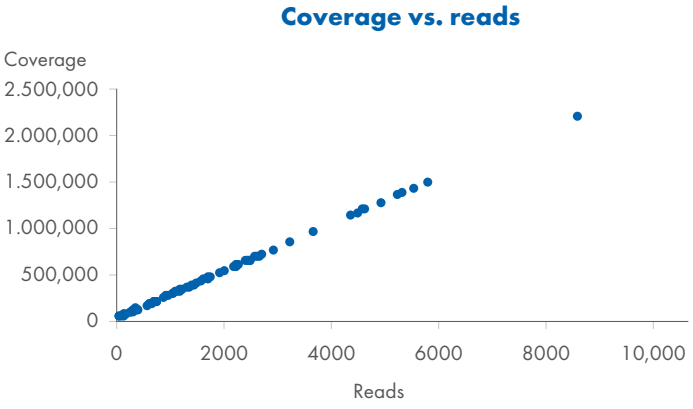

Figure 8. Sequencing data comparing the coverage against the number of reads/samples. It reinforced the expected correlation between reads received per sample and coverage across the genome.

Conclusion

High library yields and consistent coverage across the SARS-CoV-2 genome were obtained using the QIAseq DIRECT Enhanced SARS-CoV-2 library prep workflow on the Sciclone NGSx iQ Workstation. Sequencing data showed successful targeting of the SAR-CoV-2 with high mapping to the genome. Moreover, the automated workflow produced higher quantities of final libraries than the manual workflow. Automating complex and labor-intensive NGS workflows can greatly increase throughput while decreasing human error and maintaining accuracy. This solution is scalable for varying sample types, titers and batch sizes to generate high-quality sequencing libraries in low-and high-throughput conditions.

References

1. ARTIC network; **https://artic.network/**

Ordering Information

Product	Contents	Cat. no.
QIAseq DIRECT SARS-CoV-2 Kit A	Single-box solution containing all materials for reverse transcription and library prep with 96 Unique Dual Indices (Set A)	333891
QIAseq DIRECT SARS-CoV-2 Enhancer	Contains all supplemental material necessary for the Enhanced QIAseq DIRECT SARS-CoV-2 workflow	333884

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services or your local distributor.

To request QIAseq library prep automation scripts, visit: **www.qiagen.com/applications/next-generation-sequencing/dna-sequencing/automated-library-prep**