Automation of the QIAseq® DIRECT SARS-CoV-2 library and enhancer kits on the Sciclone NGSx iQ workstation.

This workflow was demonstrated in Dr. Pavitra Roychoudhury's laboratory at University of Washington, Department of Laboratory Medicine & Pathology, Seattle, WA, USA with the support from QIAGEN and Revvity.

Sciclone NGSx iQ workstation

Introduction

The SARS-CoV-2 virus is the causative agent of the ongoing COVID-19 pandemic. With the fast-spreading nature of the virus, 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) have swiftly emerged. Next-generation sequencing (NGS) has become a vital tool in the detection of these variants. Novel variants are still being discovered and have become an area of concern that requires sequencing to identify their epidemiological prevalence. For example, mutations in the spike protein can cause increased transmissibility. The genome size of the SARS-CoV-2 virus is relatively small at 30kb and can be detected in a variety of sample sources such as nasopharyngeal swabs, buccal swabs, and wastewater.

There are unbiased meta-transcriptomics approaches to recover viral, host, and bacterial gene expression by aligning against viral and host transcriptomes. Alternatively, amplicon-based capture strategies of the SARS-CoV-2 virus can also provide high sensitivity from limited amounts of total RNA. Amplicon sequencing allows for sequencing only the SARS-CoV-2 virus without wasting reads on other RNAs. Typical sequencing of the virus is achieved by isolation of RNA, conversion of RNA to cDNA, enrichment of the SARS-CoV-2 virus, fragmentation of target enriched libraries, adapter ligation, library amplification, and sequencing. Since the beginning of the pandemic, most primer designs have been based on primer sequences from the ARTIC network, amplifying 400 bp amplicons into two PCR pools that cover 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 QIAGEN® Enhanced QIAseq® DIRECT SARS-CoV-2 protocol introduces a streamlined, four-hour workflow for the enrichment and library prep. It removes the need for fragmentation following target enrichment and does not need quantification/normalization before library amplification and indexing. This reduces library preparation time and allows for more samples to be multiplexed on one sequencing run. It maintains robustness, uniformity of coverage, and allows for better variant discovery compared to the ARTIC network-based protocols.

The process of library preparation can be a pain point for operators spending long periods at the bench and waiting for long incubation times. Automation of library preparation protocols can decrease operational cost, error rate, and sample variability. Steps such as magnetic bead washing to remove remains of PCR reactions specifically can be the source of low overall library yields. Consistency of bead washing can result in higher library yields. Processing large batches of samples manually can lead to more opportunities for user error, and by elimination of human touchpoints, automated workflows can decrease the number of reruns of libraries due to preventable operator error. Automation also allows an operator to reliably process small or large batches of samples at a time. The Sciclone® G3 NGSx iQ™ has an on-deck thermocycler and tips that allow an operator to start the program and allow the liquid handler to proceed with the 6-hour workflow. The instrument is also capable of pipetting small volumes often used in NGS workflows. We describe the detection of SARS-CoV-2 variants using QIAseq® DIRECT protocol on the Illumina® NextSeq 2000® sequencer automated on the Sciclone G3 NGSx iQ workstation allowing a reliable, high throughput process to detect variants of the SARS-CoV-2.

Methods

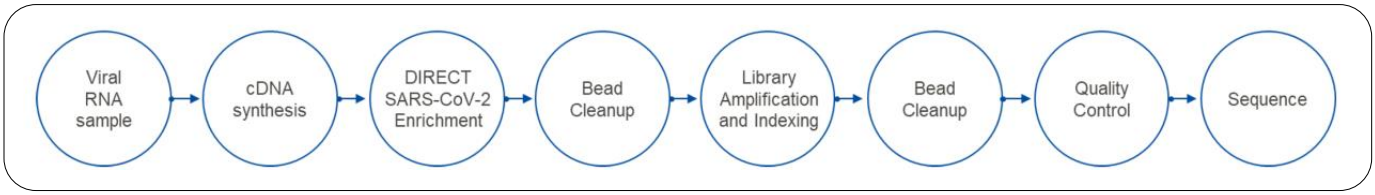
Library preparation

1. Viral RNA samples
2. cDNA synthesis (75 minutes automated) (75 minutes manual)
3. Direct SAR-CoV-2 Enrichment (130 minutes automated) (145 minutes manual)
4. Bead Cleanup (40 minutes automated) (45 minutes manual)
5. Library Amplification and Indexing (45 minutes automated) (45 minutes manual)
6. Bead Clean up (40 minutes automated) (45 minutes manual)
7. QC
8. Sequencing

Total Time: 5.5 hours automated 6hrs manual

Viral RNA was extracted using 400 µL of each specimen eluted to 50 µL on the KingFisher™ Flex Purification System (Thermo Fisher, Waltham, MA) following the manufacturer's instructions. Then enriched sequencing libraries were generated on the Sciclone G3 NGSx iQ workstation in a high throughput method. The Sciclone workstation has prompts in the User Interface that allow a user to select the number of columns to process from 8 up to 96 samples at a time and review the correct deck setup for the workflow (Figure 2). The preparation of master mixes and reagents are according to the excel workbook which calculates the necessary reagents and plate/tubes necessary for the method (**Figure 1**).

The QIAseq® DIRECT SARS-CoV-2 library preparation protocol begins with 5 µL of extracted RNA which was transcribed into complementary cDNA. A total of 16 µL from the cDNA reaction is prepared with two



pools of approximately 225–275 bp QIAseq® DIRECT SARS-CoV-2 overlapping amplicons, the pools are combined then 24 µL of the target enrichment reaction regardless of viral titer is amplified and sample-indexed. The quality and quantity of libraries were assessed on a subset of samples and 5 µL of each library was pooled and sequenced on an Illumina® Nextseq2000® 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.

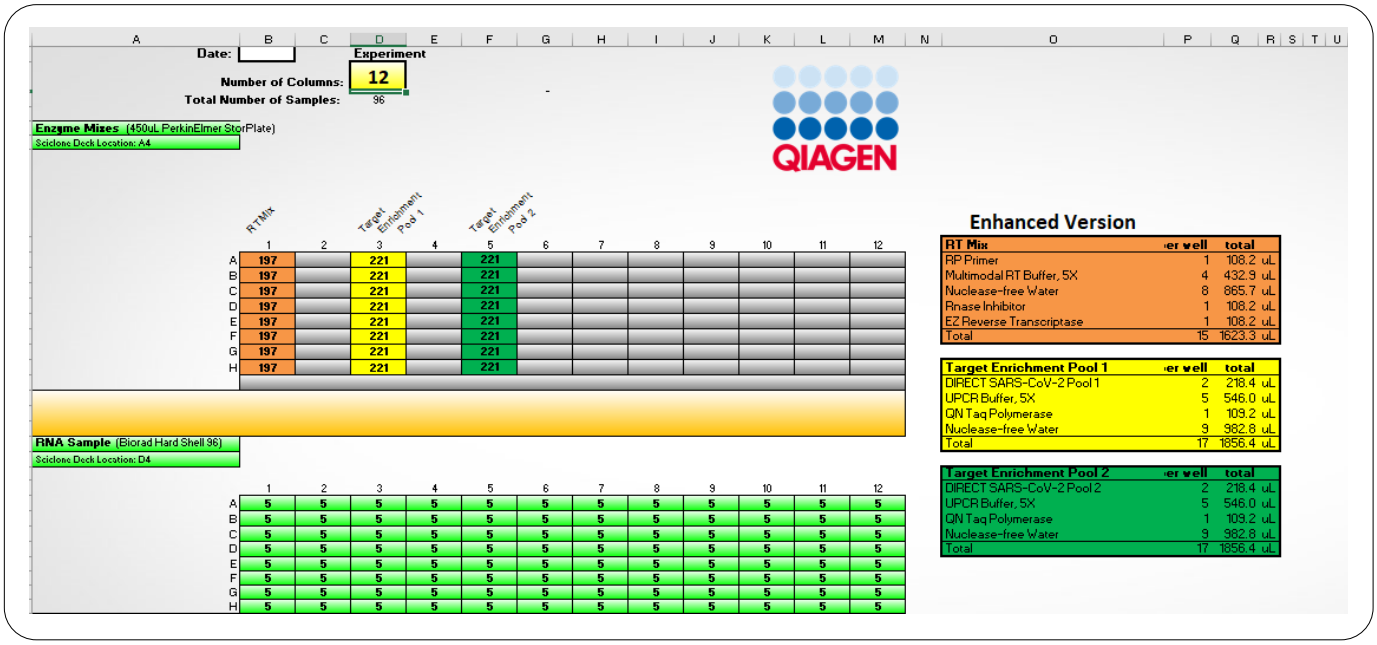


Figure 1: An example of the excel workbook describing the recipes and volumes required for different master mixes. The number of columns can be adjusted to support a variety of sample quantities throughout the steps within the workflow, for the QIAseq Direct enhanced protocol on the Sciclone G3 NGSx iQ workstation.

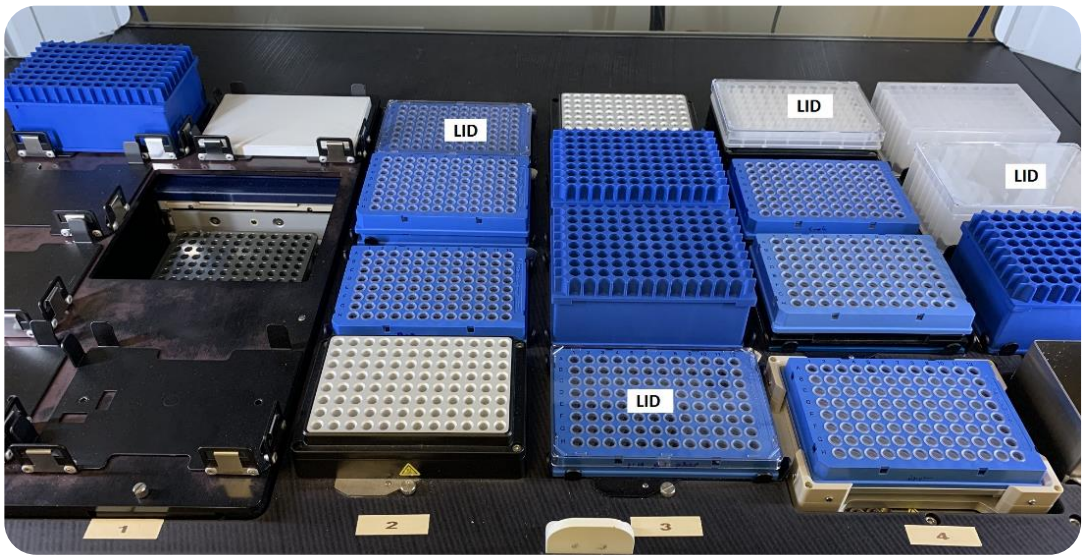


Figure 2: Example of the deck layout for the QIAseq Direct Enhanced protocol setup during reverse transcription – the first step of the workflow on the Sciclone G3 NGSx iQ workstation.

Results

Here we show successful and reliable automation of the enhanced QIAseq® Direct Library Kit for the Sciclone NGSx iQ Workstation. Comparison of the manually prepared libraries shows concurrence with the high-throughput automated workflow. The library yield was higher in libraries generated with the Sciclone workstation as compared to the manually generated libraries as shown in **Figures 4 and 5** with the overall median concentration being nearly double for the region molarity and median concentration. **Figures 9 and 10** show examples of the final library show strong libraires at about 399 bp for the automated workflow. Then manual workflow libraries (**Figures 11 & 12**) show libraires that varied having some of the major peaks at 400 bp but also seeing primer dimer at ~180 bp. This shows the concurrence in library generation for the automated workflow when compared to the manual. This process is efficient and reproduceable with a wide variety of sample types and titers. **Figure 6** illustrates that with Ct values varying on the automated run from 17-35, the coverage far exceeded 20X. This shows that there is no need for batching samples based on Ct value; Samples with a wide range of Ct can be multiplexed on one run providing acceptable sequencing data with >80% of total reads mapping to the SARS-CoV-2 genome (**Figure 3**).

Commonly acceptable data for SARS-CoV-2 sequencing requires greater than 20X coverage and traditionally the spike protein has been challenging to achieve high coverage. As displayed in **figure 8** only one sample was unable to achieve more than 20% of the reads gaining 100X coverage, which far exceeds the 20X coverage standard. The median number of reads for the run was 321,810 (min. 31176, max. 2197842). **Figure 7** depicts the correlation between coverage and reads. It shows a positive correlation that with more reads/samples the better overall coverage of the genome. 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 even with significantly lower number of reads and varying viral titer, reproducible data can be generated.

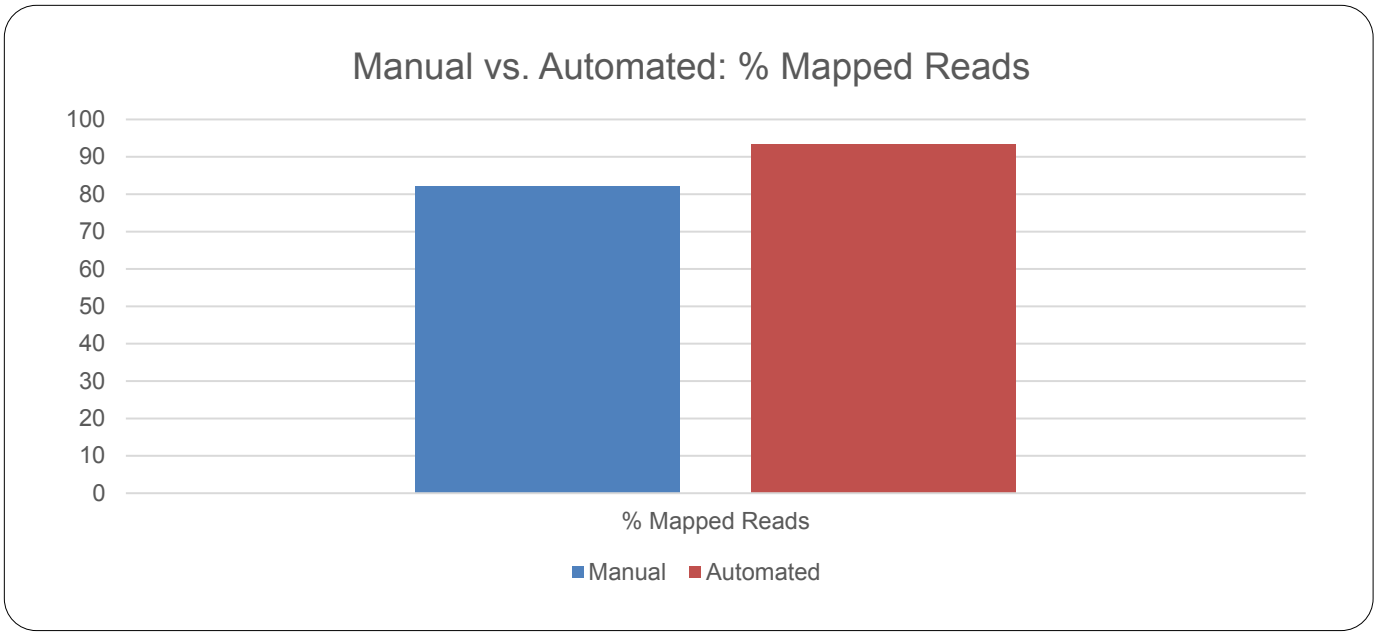


Figure 3: Sequence data measuring the percentage of reads mapped to the SAR-CoV-2 genome

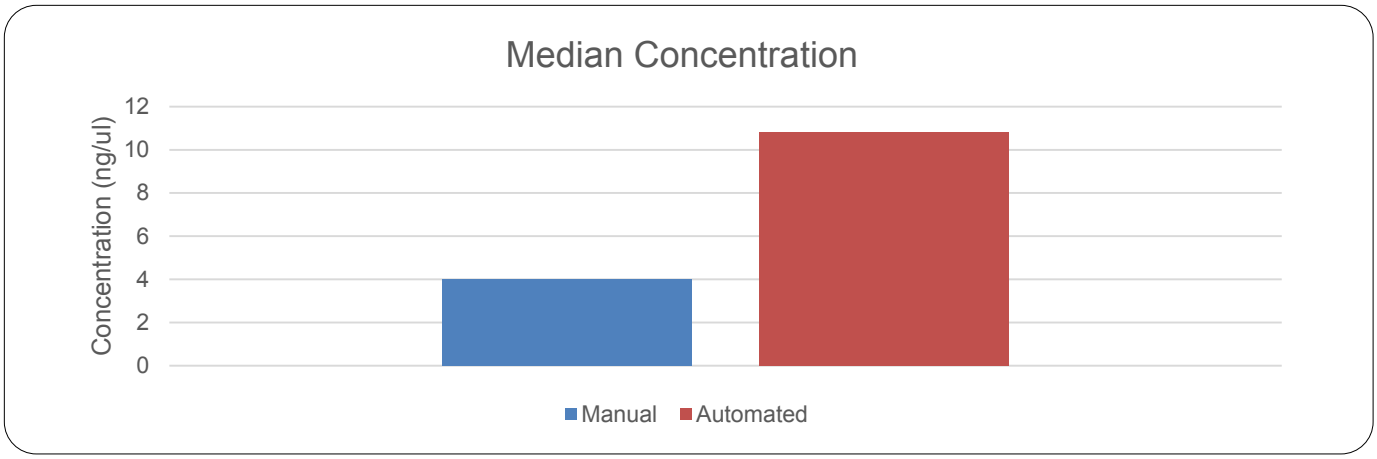


Figure 4: Comparison of median library concentration (ng/µL) following library preparation between manual and automated workflows.

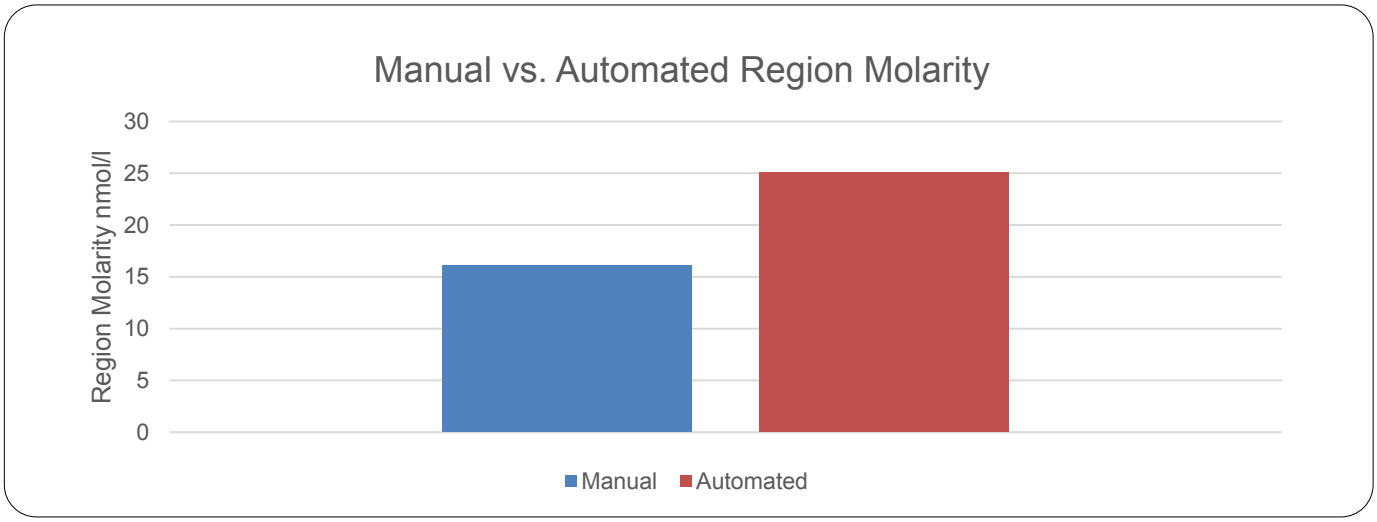


Figure 5: Region molarity (nmol/L) comparisons of final libraries for manual and automated workflows. The expected library size was approximately 400bp and the region molarity was measured between 270-750bp.

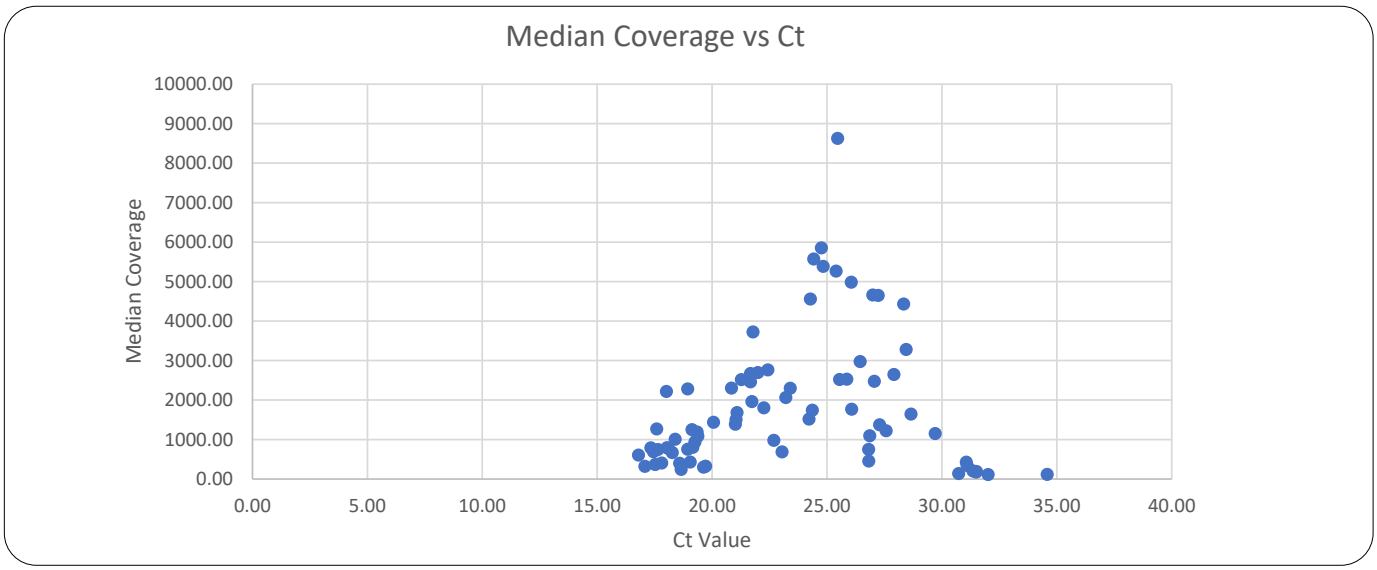


Figure 6: Sequence data mapping the coverage vs. the Ct of the input samples prior to cDNA synthesis. This shows coverage greater than 20X for all samples despite varying initial Ct values ranging from ~17-35.

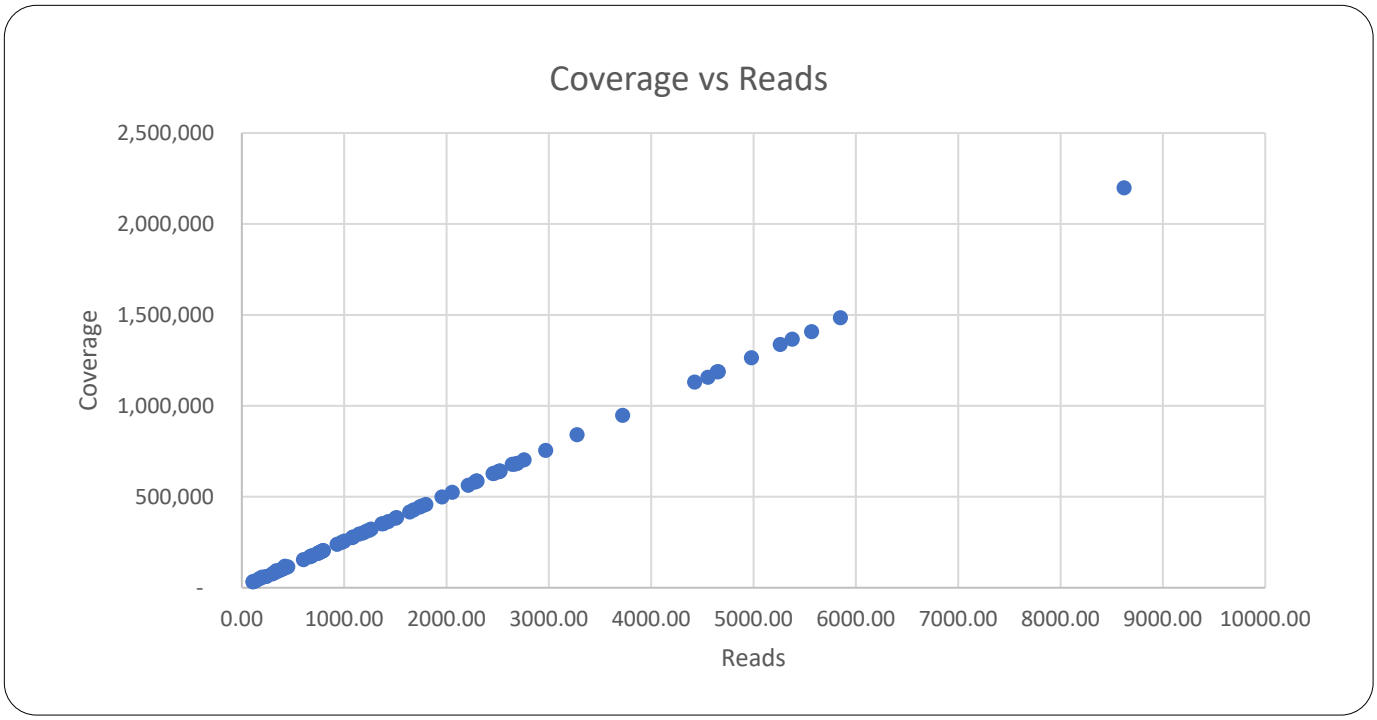


Figure 7: Sequencing data comparing the coverage against the number of reads/samples. It reinforced the expected correlation between reads received per sample and coverage across the genome.

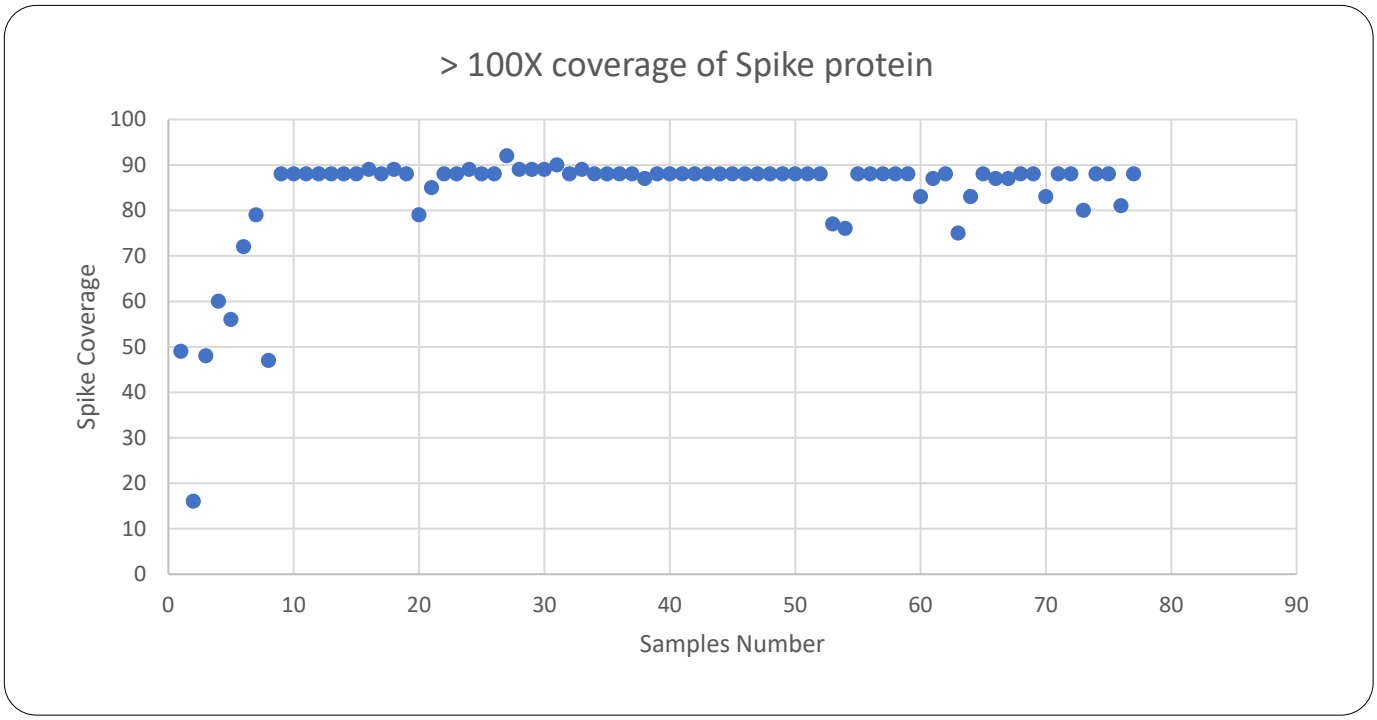


Figure 8: Sequencing data measuring the percentage of the reads mapping to the spike protein that are greater than 100X coverage. The general standard for acceptable coverage of sequencing of the SARS-CoV-2 virus is measured by the % of reads having >20X coverage. This depiction shows coverage far exceeding that standard.

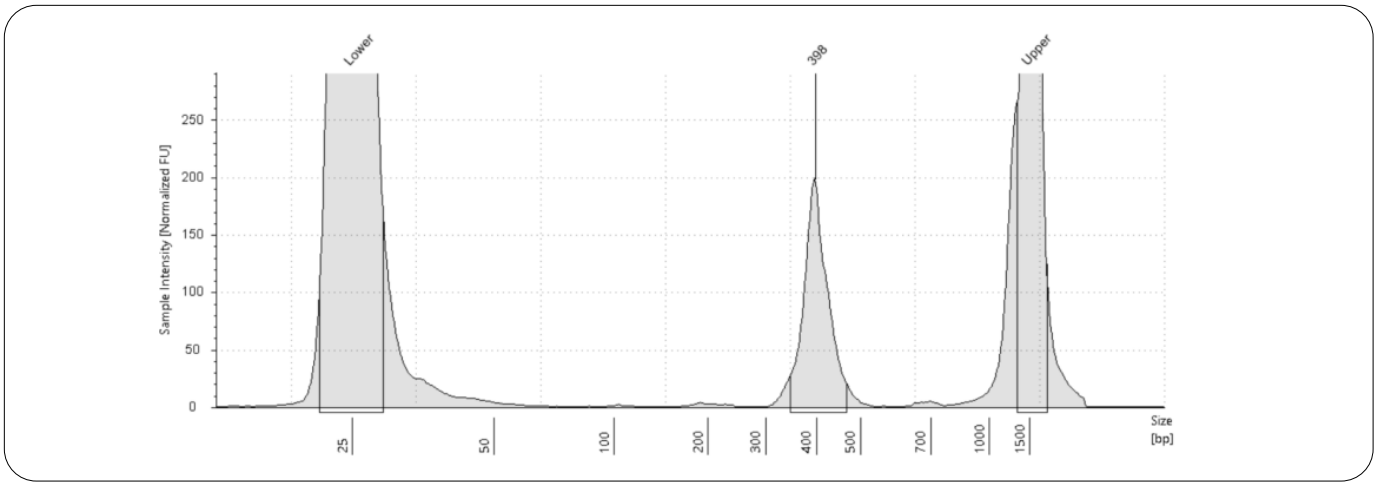


Figure 9: Library Traces, Twist Control-Automated

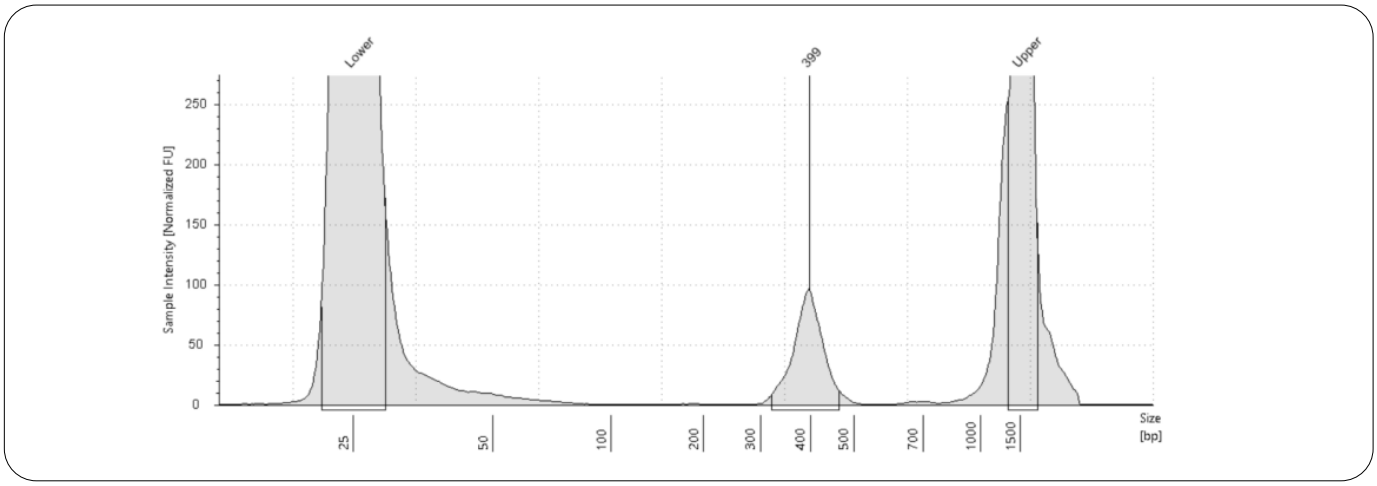


Figure 10: Library Traces Twist Control-Automated

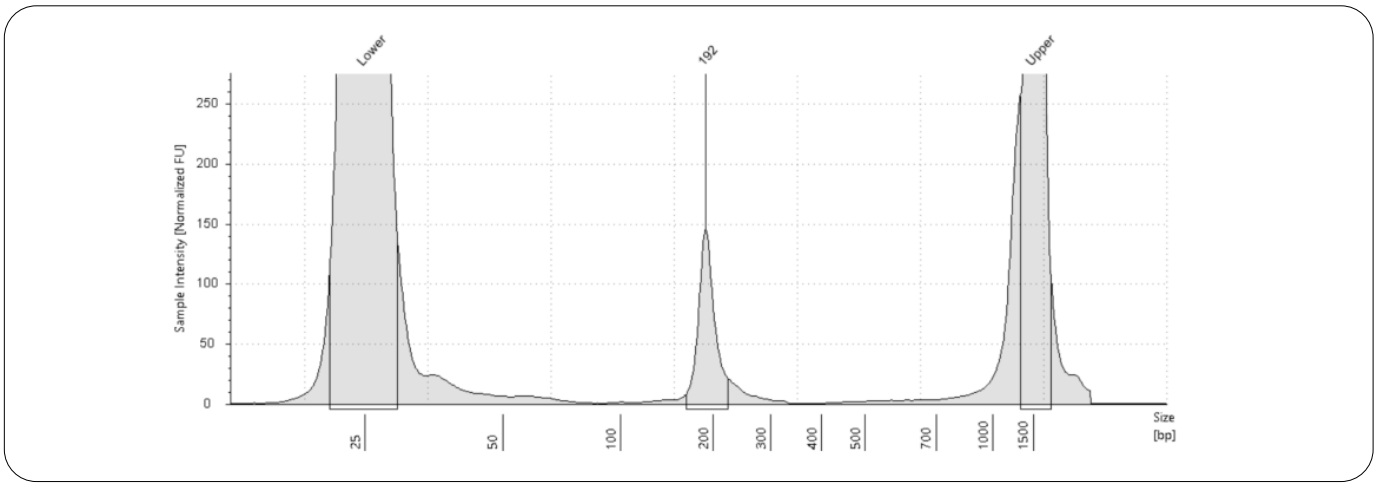


Figure 11: Library Traces, Twist Control-Manual

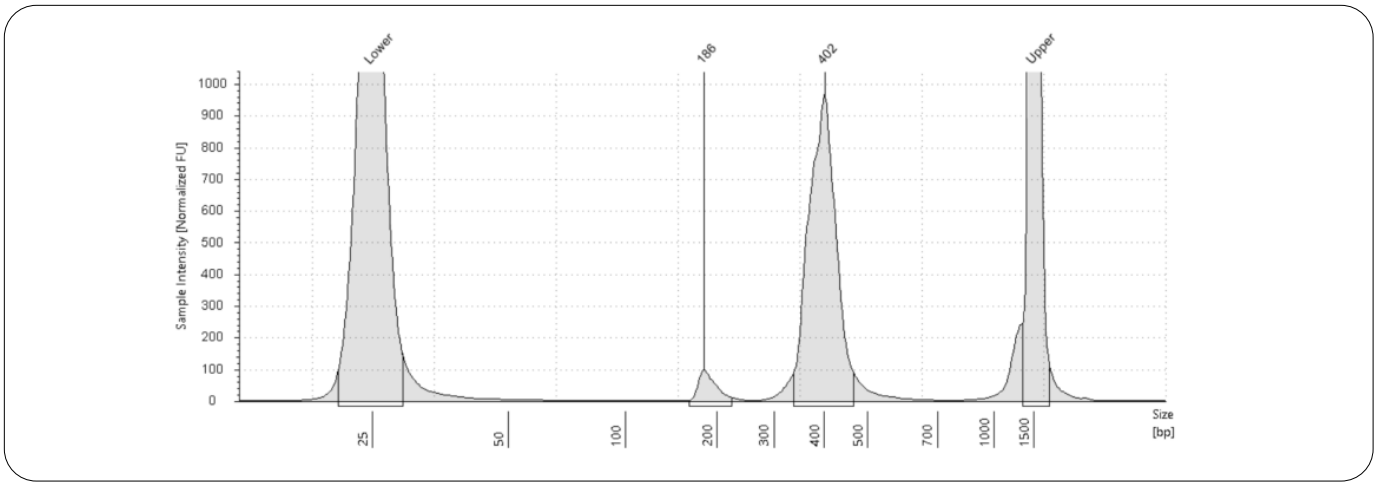


Figure 12: Library Traces, Twist Control-Manual

Conclusion

High library yields and consistent coverage across the SARS-CoV-2 genome were obtained. Sequencing data showed successful targeting of the SAR-CoV-2 with high mapping to the genome. Automation of complex and labor-intensive NGS workflows can greatly increase throughput and maintain quality standards while decreasing hands-on time and risk of human error. The automated workflow produced higher quantities of final libraries when compared to the manual workflow time. This solution is scalable for varying sample types, titers, and batch sizes to generate valued sequencing libraries in both low-and high-throughput conditions.

Revvity, Inc.
940 Winter Street
Waltham, MA 02451 USA

(800) 762-4000
www.revvity.com