APPLICATION NOTE

www.swiftbiosci.com

TARGETED SEQUENCING OF SARS-COV-2: SWIFT RNA LIBRARY KIT AND ARBOR BIOSCIENCES myBAITS EXPERT VIRUS PANEL

SUMMARY

The emergence of SARS-CoV-2 has resulted in the need for cost-effective and rapid strategies to characterize the viral genome. Complete genome sequencing is critical for identifying mutational variants that will impact research in the fields of virology, immunology, epidemiology, and molecular evolution, among others. However, maximizing viral RNA sequencing reads from clinical specimens or laboratory samples can be challenging, particularly due to the amount of host background that may be present. Here, we present a targeted next-generation sequencing (NGS) strategy that combines the Swift RNA Library Kit with the Arbor Biosciences myBaits Expert Virus SARS-CoV-2 panel. This approach has the following benefits:

- Detection down to 10 viral genome copies in a background of 50 ng human gRNA
- >99.9% coverage of the SARS-CoV-2 full-length genome
- Enrichment rates over 100,000-fold
- >100X genome coverage depth with fewer than 100,000 paired-end reads
- Multiplex up to 384 samples in a single MiSeq® run with >50X coverage per sample

INTRODUCTION

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, emerged into the human population in December 2019. SARS-CoV-2 is a positive-sense, single-stranded RNA (ssRNA) virus with a genome size of 29.9 Kb. Coronavirus particles are made up of an envelope with protruding spike proteins that surrounds and protects the viral RNA genome. Inside the envelope, the genome is bound to nucleocapsid proteins, which are important for both virus replication and assembly[a].

Rapid detection methods for SARS-CoV-2 include RNA extraction followed by reverse transcription to convert the viral RNA genome copies into complementary DNA (cDNA). The cDNA can then be used as template for qPCR, which targets and amplifies a small portion of the viral genome to confirm the presence or absence of that targeted sequence. At present, approved qPCR diagnostic tests target a variety of regions, including the nucleocapsid[b] and envelope genes, among others.

In addition to the rapid detection of SARS-CoV-2, it is important to characterize the complete genome. RNA viruses mutate rapidly and the ability to track these mutations has applications across multiple fields of biology[c]. Current efforts to track the genetic diversity of SARS-CoV-2 have resulted in an enhanced understanding of virus transmission within and between populations through phylogenetic analyses of the complete genome sequences[d]. Future studies can focus on identifying mutations that may impact viral pathogenesis; these studies will require increased sequencing of patient samples and correlation of sequence data with symptoms and disease outcomes.

In order to improve SARS-CoV-2 complete genome sequencing efforts, we present a targeted NGS workflow. Coronaviruses have the largest genomes of all RNA viruses. However, while a 29.9 Kb genome is considered large for an RNA virus, it is highly tractable for a targeted NGS approach. For most SARS-CoV-2 patient samples, host RNA will comprise the majority of the sample. This makes direct sequencing of an RNA-Seq library prepared from patient samples inefficient. However, the use of a hybridization capture panel specific to the SARS-CoV-2 genome can enrich for viral genome sequences and dramatically reduce sequencing costs. We pair the Swift RNA Library Kit[e] with the Arbor Biosciences myBaits Expert Virus SARS-CoV-2 Panel[f] to enrich for and sequence full-length SARS-CoV-2 genomes from a mixed RNA sample (**Figure 1**).

SWIFT RNA LIBRARY KIT

The Swift RNA Library Kit (Cat. Nos. R1024, R1096) leverages Swift Biosciences' patented Adaptase® technology, a single-stranded tailing and ligation reaction, to directly convert first-strand cDNA into NGS libraries. The use of Adaptase in the Swift RNA workflow offers several advantages: i) library molecules maintain strandedness because a functional library molecule is only produced from first-strand cDNA; ii) the simplified workflow does not require conventional second-strand cDNA synthesis; iii) ligation efficiency is maintained at low inputs, eliminating the need for adapter titration; and iv) the Adaptase and Ligation reactions are fully separated, resulting in substantially reduced adapter dimers in the final libraries.

The Swift RNA Library Kit is a fast, robust workflow that can be completed in ~4.5 hours and can support down to 100 pg of RNA as direct input (i.e., no upstream processing). For traditional transcriptomic approaches, the Swift RNA kit results in high mapping rates, with a superior number of genes and transcripts detected, as well as even coverage across all transcripts. For hybridization capture applications, the Swift RNA kit produces generous yields and performs well even with damaged or degraded samples (see the Application Note: The Swift RNA Library Kit Optimizes RNA-Seq Data and Costs for FFPE Samples). The efficient conversion of RNA into library molecules from low inputs and challenging samples makes the Swift RNA Library Kit an ideal choice for working with viral RNA samples.

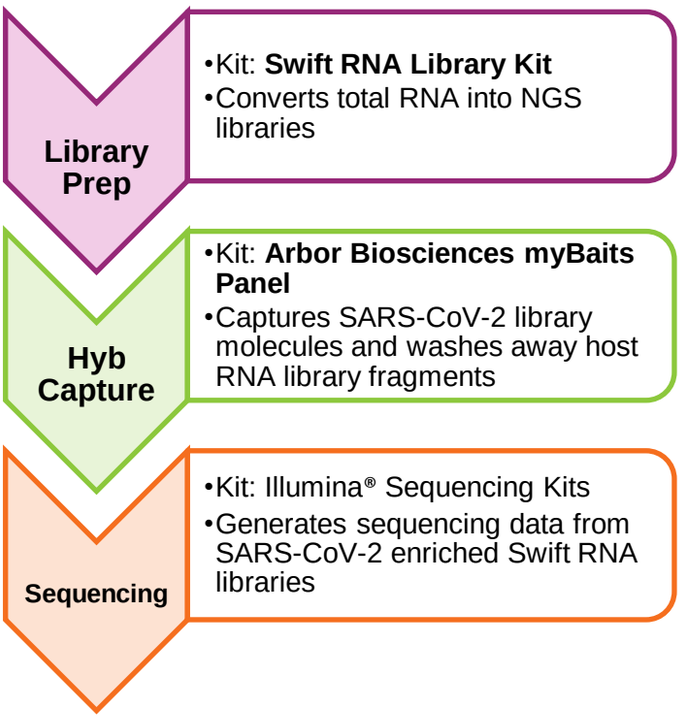

Figure 1: SARS-CoV-2 Targeted NGS Workflow: combine the Swift RNA Library Kit with the Arbor Bioscience myBaits SARS-CoV-2 Panel.

myBAITS EXPERT SARS-CoV-2 PANEL

The myBaits Expert Virus SARS-CoV-2 Panel (Cat. Nos. 308716, 308748, 208796) comprises over 2,000 biotinylated ssRNA probes. These probes are 80 nucleotides in length and were designed using all full and partial SARS-CoV-2 genome sequences available in NCBI as of January 31, 2020. Hybridization capture works by capturing and enriching for library molecules that have sequences complementary to the designed probes. An RNA sample containing host and viral RNA can be processed through the Swift RNA kit to produce fully functional library molecules from all RNA species within the mixed sample. The myBaits hybridization capture workflow would then enrich for just the library molecules that originated from SARS-CoV-2 RNA.

Hybridization capture requires only a single, short hybridization site to successfully capture a molecule. This allows the system to retrieve viral RNA fragments of a wide range of lengths, making the myBaits workflow appropriate for targeted sequencing of both fresh viral isolates and highly degraded trace RNA sources, as well as samples with high levels of genetic diversity. Finally, because hybridization capture is typically performed on individually-barcoded sequencing libraries, multiple samples can be pooled per reaction, offering tremendous scalability benefits.

METHODS

RNA libraries were made using the Swift RNA Library Kit (Cat. No. R1096). Either 10 or 50 ng of Universal Human Reference (UHR) RNA (Agilent 740000) was used as non-target background and combined with varying numbers of SARS-CoV-2 synthetic viral genome copies (Twist Biosciences 102024). The SARS-CoV-2 substrate is present as 6 x 5 kb ssRNA fragments representing an early Wuhan isolate (MN908947, NC_045512). Spike-ins of 0, 10, 10^3, or 10^6 viral genome copies were tested in duplicate. Briefly, the UHR + SARS-CoV-2 mixed RNA samples were used as input into the Swift RNA Library Kit, with protocol adjustments specific to libraries intended for downstream hybridization capture: fragmentation times were increased to 15 min; SPRI ratios throughout the protocol were increased to 1.8X; and the number of PCR cycles was increased (15 cycles for 10 ng input, 13 cycles for 50 ng input). Libraries were

indexed using the Swift Combinatorial Dual Indexing Primer Kit (Cat. No. X8096), quantified with a Qubit dsDNA HS assay (Cat. No. Q32851), and evaluated on an Agilent Bioanalyzer HS DNA Chip (Cat. No. 5067-4626). Libraries containing 10^6 SARS-CoV-2 genome copies were sequenced prior to hybridization capture to assess SARS-CoV-2 content using 2 x 76 bp High Output chemistry with dual 8-nt indexing reads (Cat. No. FC-420-1002) on an Illumina MiniSeq®.

Hybridization capture was performed using the Arbor Biosciences myBaits Expert Virus SARS-CoV-2 panel (Cat. Nos. 308716, 308748, 208796). Each myBaits target enrichment reaction contained 200 ng of 4 individual libraries for a total of 800 ng per capture pool. Libraries of comparable starting SARS-CoV-2 genome copies were pooled together to minimize sample dropout effects. The procedure followed the default myBaits version 4.0 protocol[g] but used 0.5 µL IDT xGen Blocking Oligos TS mix (Cat. No. 1075474) in lieu of Block A reagent. Following capture clean-up, bead-bound enriched libraries were amplified with universal P5/P7 primers for 14 cycles using the KAPA HiFi polymerase system (Cat. No. KK2601), purified with SPRI beads, and quantified using the KAPA Illumina library qPCR system (Cat. No. KK4824). Enriched libraries were pooled in equimolar ratios and sequenced on an Illumina MiSeq® using a 2 x 76 bp protocol with dual 8-nt indexing reads (Cat. No. MS-102-3001).

Demultiplexed FASTQ data were aligned with bwa mem (version 0.7.5a-r405)[h] to a reference file containing both the complete GRCh38 human reference genome with "ALT" entries removed, and a SARS-CoV-2 genome assembly (GenBank Acc. No. NC_045512.2). Alignments were converted to sorted BAM files with samtools (version 0.1.19-96b5f2294a)[i]. Coverage analysis was performed following deduplication with picard MarkDuplicates (version 2.18.15)[j]

RESULTS

The myBaits SARS-CoV-2 panel is capable of fully enriching SARS-CoV-2 library molecules within a mixed sample of SARS-CoV-2 and host background fragments. Swift RNA libraries made from 10^6 SARS-CoV-2 genome copies in a background of either 10 or 50 ng UHR RNA were sequenced pre- and post-hybridization capture. Without hybridization capture, SARS-CoV-2 reads make up less than 0.1% of the sequencing data. However, after hybridization capture, the percent of reads mapping to the SARS-CoV-2 genome increases to >85% (**Table 1**). This demonstrates a several thousand-fold enrichment with just a single round of hybridization capture.

In addition to substantially enriching SARS-CoV-2 library molecules, the myBaits panel also preserves the comprehensive SARS-CoV-2 genome representation within the Swift RNA libraries. Comparing the coverage plots of libraries sequenced pre- and post-hybridization capture shows very similar coverage profiles (**Figure 2**). Importantly, virtually no gaps in coverage were observed when using the Swift RNA kit and myBaits capture (**Figure 2**). More than 99.9% of the genome shows >10X coverage for samples with 10^6 viral genome copies at a sequencing depth of 250,000 reads (**Figure 3**, left panel). Coverage of the complete SARS-CoV-2 genome is important because it enables detection of all possible nucleotide variants; tracking these mutations is crucial for accurate phylogenetic inferences and epidemiological studies. At a depth of 250,000 reads per library, the Swift RNA + myBaits workflow can detect as few as 10^3 viral genome copies in a mixed starting sample and cover ~80% of the full-length genome. Even with only 10 viral copies present in the starting RNA sample, sequencing reads were retrieved across ~15% of the viral genome (**Figure 3**, left panel), demonstrating the ability to detect even trace amounts of viral RNA.

The 29.9 Kb SARS-CoV-2 genome is a relatively small target region. This enables many samples to be multiplexed in a single sequencing run to provide efficient and economical characterization of the viral genome.

UHR (ng)	SARS-CoV-2 (copies, pg)	Expected SARS-CoV-2 Representation (%)	Reads mapping to SARS-CoV-2 (%) pre-hyb capture	Reads mapping to SARS-CoV-2 post-hyb capture (%)	Fold enrichment (X)
10	10^6 (20 pg)	0.20	0.04	97.4	2,435
50	10^6 (20 pg)	0.04	0.01	86.2	8,620

Table 1. Data metrics for SARS-CoV-2 libraries sequenced pre- and post-hybridization capture. Libraries were made with the Swift RNA kit and captured with the Arbor Biosciences myBaits SARS-CoV-2 panel.

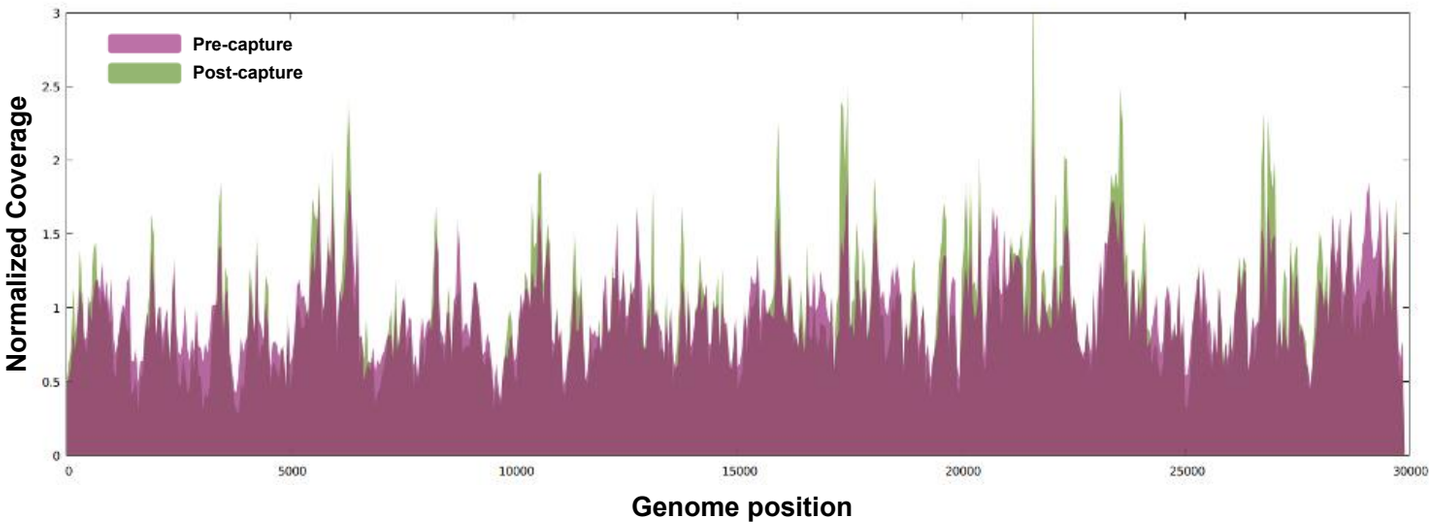

Figure 2: Normalized coverage of the SARS-CoV-2 genome pre- and post- hybridization capture (purple and green, respectively). The similarity of the coverage profiles demonstrates the ability of the myBaits panel to preserve the comprehensive viral genome representation in the starting Swift RNA library.

In order to understand how sequencing depth impacts the mean base coverage along the SARS-CoV-2 genome, we downsampled sequencing data from 500,000 down to 10,000 reads per library. When samples have a starting viral load of 10^6 genome copies, ~25X coverage of the full-length genome can be achieved with just 10,000 reads (**Figure 3**, right panel). Multiplexing 384 samples in a single MiSeq sequencing run (~65,000 reads per sample) yields >150X mean coverage with 10^6 starting viral genome copies. This is more than sufficient for confident nucleotide variant calling, while ~2X mean coverage with 10^3 starting viral copies is sufficient for SARS-CoV-2 detection. With just 250,000 reads, the standard myBaits workflow can detect virus down to 10 viral copies. Broader viral genome characterization is possible with a starting input of 10^3 viral copies (**Table 2**). Enrichment rates range from 435-fold (10^6 viral copies) to over 100,000-fold (10 viral copies), demonstrating the

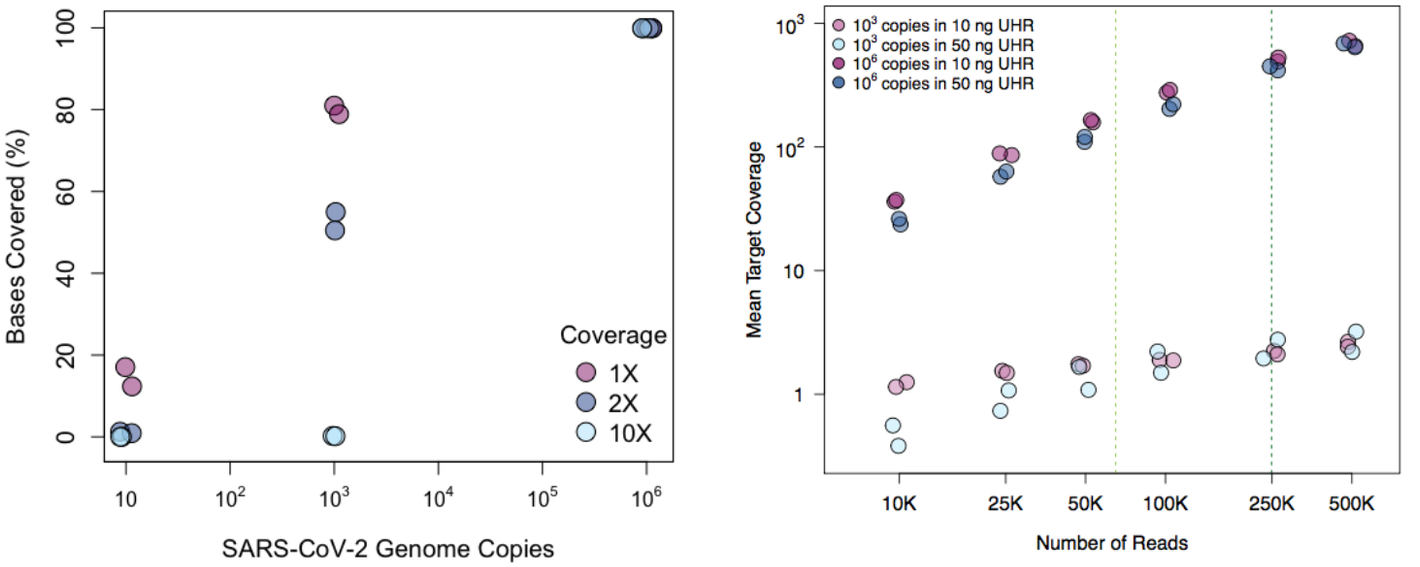

Figure 3: Coverage of the SARS-CoV-2 Genome. Left panel: Percent genomic bases covered vs. the number of SARS-CoV-2 genome copies spiked into 10 ng UHR. Libraries were sequenced to a depth of 250,000 reads per sample. Right panel: Mean target coverage of the SARS-CoV-2 genome vs. the number of paired-end reads per sample for libraries made with either 10^3 or 10^6 viral genome copies spiked into either 10 or 50 ng UHR. The light and dark green dashed lines represent the number of reads permitting 384 or 96 samples to be multiplexed on a single 2 x 76 bp MiSeq run.

ability of the capture panel to select and enrich for SARS-CoV-2 library molecules (**Table 2**). Additional strategies are possible for enrichment of samples with low starting viral titers, including potential modifications to both library preparation (e.g., by increasing the total RNA mass used in the library preparation, and/or performing a ribosomal RNA depletion step to increase the viral to host RNA ratio) and/or target enrichment conditions (e.g. by increasing mass input to hybridization, reducing hybridization and wash temperatures, and/or performing two rounds of enrichment).

UHR (ng)	SARS-CoV-2 (copies)	Reads mapping to SARS-CoV-2 (%)	Fold Enrichment (X)	Bases Covered 1X (%)	Bases Covered 2X (%)
10	0	0.01	-	4.1	0.0
10	10	0.04	20,000	14.7	1.1
10	10^3	6.43	32,150	79.9	52.7
10	10^6	87.07	435	99.9	99.9
50	0	0.02	-	5.0	0.2
50	10	0.05	127,500	20.8	5.2
50	10^3	1.14	28,500	81.7	54.4
50	10^6	58.21	1,455	99.9	99.9

Table 2. Data metrics for SARS-CoV-2 libraries sequenced post-hybridization capture. Libraries were made with the Swift RNA kit and enriched for SARS-CoV-2 using the Arbor Biosciences myBaits panel, sequenced on a MiSeq with 2 x 76 bp, and downsampled to 250,000 reads. Values represent the mean from two replicate libraries.

CONCLUSIONS

Sequencing the complete SARS-CoV-2 genome is important for confidently detecting nucleotide variants that can provide insight to virus evolution, transmission, and pathogenesis. However, there are several challenges to sequencing the viral genome, including variable sample quality/RNA integrity, low viral titers, and high host genomic background, among others. Here we present an NGS workflow solution for sequencing SARS-CoV-2: the Swift RNA Library Kit followed by hybridization capture using the Arbor Biosciences myBaits Expert Virus SARS-CoV-2 Panel.

The Swift RNA Library Kit leverages Swift's Adaptase® technology to produce high-quality libraries from limited starting material. Swift RNA performs well with damaged or degraded samples, ensuring that low-quality viral RNA samples will be converted into library molecules. With a fast, straight-forward workflow that is easily automatable, the Swift RNA kit provides numerous benefits for processing SARS-CoV-2 samples.

Sequencing of SARS-CoV-2 libraries directly is costly due to the high proportion of reads that are taken up by host background RNA. A more efficient and cost-effective workflow is to perform hybridization capture of libraries made from SARS-CoV-2 samples. The myBaits Expert Virus SARS-CoV-2 panel from Arbor Biosciences provides a targeted solution to enrich for only the SARS-CoV-2 genome. With over 2,000 probes, the myBaits panel covers the entire viral genome, showing >99.9% coverage in samples with as few as 10^6 viral genome copies. In fact, samples with a starting input of only 10 viral copies were still able to be detected, with ~15% of the genome covered at 1X with just 250,000 total reads sequenced per library. For samples with 10^6 genome copies, ~25X coverage can be achieved with as few as 10,000 reads, greatly enhancing the ability to multiplex samples and improve the throughput and cost effectiveness of this NGS workflow. Further, we demonstrate that the hybridization capture panel does not introduce coverage bias across the genome, allowing the sequencing results to reflect the comprehensive genome representation in the starting RNA-Seq library.

Hybridization capture offers several advantages over other targeted workflows. Due to the nature of hybridization capture, the probes only need to match a small region of the library molecule. This means that the system works efficiently on damaged or degraded samples with small library insert sizes and is also robust to genetic diversity that will arise within the SARS-CoV-2 population due to high viral mutation rates. Additionally, the probe set can easily be expanded to include other coronavirus strains that may be of interest, such as the four human coronaviruses that circulate on a yearly basis, causing only mild disease symptoms. Overall, pairing the Swift RNA Library Kit and Arbor Biosciences myBaits SARS-CoV-2 Panel provides a simple solution for versatile, reproducible, and cost-effective sequencing of SARS-CoV-2 samples.

ORDERING INFORMATION

Product	Reactions	Catalog Number
Swift RNA Library Kit	24	R1024
	96	R1096
Single Indexing Primer Kit Set A (12 indices, 2 rxns ea)	24	X6024
Combinatorial Dual Indexing Primer Kit (96 combinations, 1 rxn ea)	96	X8096
Set S1-S4 Combinatorial Dual Indexing Primer Kits	24 x 8 (192)	X85192 – X88192
	96 x 8 (768)	X89768
Unique Dual Indexing Primer Kits (96 UDIs)	96	X9096
	96	X9096-PLATE
	384	X90384
Arbor Biosciences myBaits Expert SARS-CoV-2 Panel	16	308716
	48	308748
	96	308796
Swift Library Amplification Primer Mix	96	88196
KAPA HiFi HotStart ReadyMix	50	KK2601
IDT xGen® Blockers TS Mix	16	1075474

REFERENCES

a) Weiss and Leibowitz. 2011. Coronavirus Pathogenesis. Advances in Virus Research. 81:85-164. https://doi.org/10.1016/B978-0-12-385885-6.00009-2

b) CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel. https://www.fda.gov/media/134922/download

c) Peck and Lauring. 2018. Complexities of viral mutation rates. Journal of Virology. 92:e01031-17. https://doi.org/10.1128/JVI.01031-17

d) Genomic epidemiology of novel coronavirus. https://nextstrain.org/ncov/global

e) Swift RNA Library Kit. Swift Biosciences. https://swiftbiosci.com/swift-rna-library-kit/

f) myBaits Expert Virus – SARS-CoV-2. Arbor Biosciences. https://arborbiosci.com/genomics/targeted-sequencing/mybaits/mybaits-expert/mybaits-expert-virus-sars-cov-2/

g) myBaits protocol – version 4 https://arborbiosci.com/wp-content/uploads/2018/04/myBaits-Manual-v4.pdf

h) bwa mem (Burrow-Wheelers Alignment Tool) http://bio-bwa.sourceforge.net/bwa.shtml

i) samtools http://www.htslib.org/doc/samtools.html

j) Picard command-line tools https://broadinstitute.github.io/picard/

Swift Biosciences, Inc.
674 S. Wagner Road, Suite 100 • Ann Arbor, MI 48103 • 734.330.2568 • www.swiftbiosci.com