Optimized protocol for sequencing the SARS-CoV-2 genome

Introduction

The novel coronavirus (SARS-CoV-2) has quickly become one of the biggest global public health concerns of the modern era. The virus can cause severe respiratory illness and exhibits a wide range of symptoms. In older individuals, the threat of disease onset can be serious, especially when underlying medical issues exist [1]. The scientific community has rallied to research this threat and better understand viral function and disease mechanisms.

Genomic sequencing of SARS-CoV-2 can enable tracking of global transmission and better understanding of viral mechanisms. Faster and more sensitive analysis of the SARS-CoV-2 genome and mutations can be realized through next-generation sequencing (NGS). NGS provides a hypothesis-free method of infectious disease detection and offers insights into both the host and pathogen.

This application note describes a workflow solution for NGS library preparation of SARS-CoV-2 samples using the **Invitrogen™ Collibri™ ES DNA Library Prep Kit for Illumina™ Systems with Unique Dual (UD) Indexes**, from purification of total RNA to library quantification and sequencing (Figure 1).

Methods

Samples

Samples were obtained with consent from the Santara Clinics Biobank in Lithuania. Each sample represents a unique donor. The presence of SARS-CoV-2 within biobank samples was confirmed by C_t values of 17–23 prior to sequencing.

Purification of total RNA

Total RNA was extracted from bronchoalveolar lavage (BAL) clinical specimens. The **Applied Biosystems™ MagMAX™ Viral/Pathogen Nucleic Acid Isolation Kit** is recommended for sample purification. After lysing BAL samples, RNA was purified according to the standard protocol.

cDNA synthesis

Each RNA sample (10 µL) was reverse-transcribed into cDNA using the **Invitrogen™ SuperScript™ IV VILO™ Master Mix** and **Thermo Scientific™ Second Strand cDNA Synthesis Kit**. The standard protocols for the kits were used.

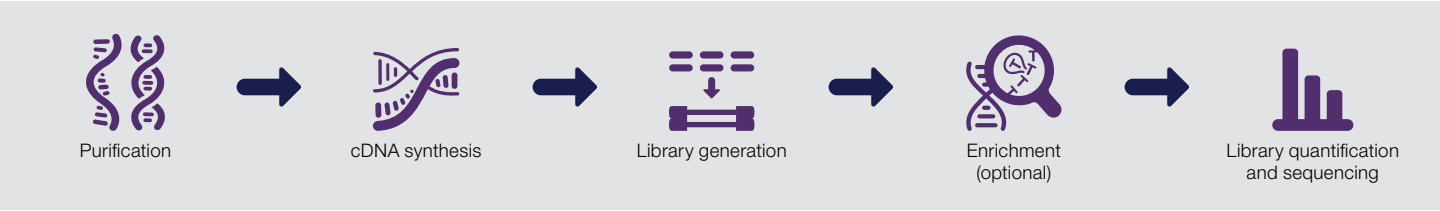


Figure 1. NGS workflow for SARS-CoV-2 samples.

Library generation

cDNA samples were converted into sequencing-ready libraries using enzymatic fragmentation with the **Collibri ES DNA Library Prep Kit for Illumina Systems with UD Indexes**. The standard protocol for the Collibri ES DNA Library Prep Kit for Illumina Systems was used with the following modifications: 25 µL of double-stranded cDNA (50 ng recommended input) was used after completing the protocol for the Second Strand cDNA Synthesis Kit (Table 1). The buffer/DNA mixture (36 µL) was prepared using 10 µL of cDNA after EDTA removal. The 5X Fragmentation and dA-Tailing Enzyme Mix (14 µL) was then added to the buffer/DNA mixture for a 50 µL total volume. Fragmentation was then performed for 20 min at 37°C. Purification of the adaptor-ligated library was accomplished according to the customized cleanup protocol (**Pub. No. MAN0019527**) with the Collibri ES DNA Library Prep Kit for Illumina Systems. The library was then amplified by PCR using 12 cycles prior to enrichment.

Enrichment (optional)

The libraries were pooled and enriched using a synthetic probe panel. Enriched libraries were amplified for 8 cycles using 2X Library Amplification Master Mix from the Collibri ES DNA Library Prep Kit for Illumina Systems (Table 2).

Table 1. Recommended protocol optimization for library generation using Collibri ES DNA Library Prep Kits.

Step	Standard recommendation			Recommended changes for SARS-CoV-2 samples	
1. Remove EDTA from DNA samples	(If needed)			Begin with 25 µL after completing reverse transcription	
	Input: 1–500 ng			Input: 50 ng	
2. Fragment the DNA and add dA tails	On ice or a cooling rack, assemble the fragmentation and dA-tailing reaction for each DNA sample in a sterile 0.2 mL thin-wall PCR tube. Add the reagents in the order given.			On ice or a cooling rack, assemble the fragmentation and dA-tailing reaction for each DNA sample in a sterile 0.2 mL thin-wall PCR tube. Add the reagents in the order given.	
	Component	**Volume**		**Component**	**Volume**
	10 mM Tris-HCl, pH 7.5–8.5	to 40 µL		cDNA	10 µL
	Double-stranded DNA (1–500 ng)	X µL		10 mM Tris-HCl, pH 7.5–8.5	to 31 µL
	10X Fragmentation and dA-Tailing Buffer (blue)	5 µL		10X Fragmentation and dA-Tailing Buffer (blue)	5 µL
	Total volume (light blue mixture)	**40 µL**		**Total volume (light blue mixture)**	**36 µL**
	Add 5X Fragmentation and dA-Tailing Enzyme Mix to the sample.			Add 5X Fragmentation and dA-Tailing Enzyme Mix to the sample.	
	Component	**Volume**		**Component**	**Volume**
	Buffer/DNA mixture from step 1 (light blue mixture)	40 µL		Buffer/DNA mixture from step 1 (light blue mixture)	36 µL
	5X Fragmentation and dA-Tailing Enzyme Mix (clear)	10 µL		5X Fragmentation and dA-Tailing Enzyme Mix (clear)	14 µL
	Total volume (light blue mixture)	**50 µL**		**Total volume (light blue mixture)**	**50 µL**
	Recommended fragmentation time and optimization range to obtain the desired fragment size			Fragment for 20 min at 37°	
		Fragmentation time at 37°C			
	Fragment size	**Recommendation**	**Optimization range**		
	150–300 bp	20 min	20–30 min		
	300–500 bp	10 min	10–20 min		
	500–700 bp	5 min	5–10 min		
3. Carry out post-ligation double-sided size selection	Double-sided size selection to match target insert size			Customized cleanup protocol (**Pub. No. MAN0019527**)	
4. PCR-amplify the library	The number of PCR cycles depends on the starting amount of DNA (i.e., input DNA).			Amplify the library for 12 PCR cycles.	

Table 2. Recommended PCR conditions to amplify enriched NGS libraries.

Stage	Number of cycles	Temperature	Time
Activate the enzyme	1 cycle	98°C	30 sec
Denature	3–4 cycles for 100 ng of input DNA 6–8 cycles of 10 ng of input DNA 10–12 cycles for 1 ng of input DNA	98°C	15 sec
Anneal		60°C	30 sec
Extend		72°C	30 sec
Final extension	1 cycle	72°C	1 min
Hold	1 cycle	4°C	Hold

Depending on the specific research goal, enrichment for SARS-CoV-2 may or may not be required. There is no enrichment needed for host–pathogen interaction studies, as sequences of both are retained. Viral genomic evolution studies, however, will require enrichment following library preparation in order to enhance the percentage of reads mapping to the SARS-CoV-2 genome. This will also reduce the cost of sequencing.

Library quantification and sequencing

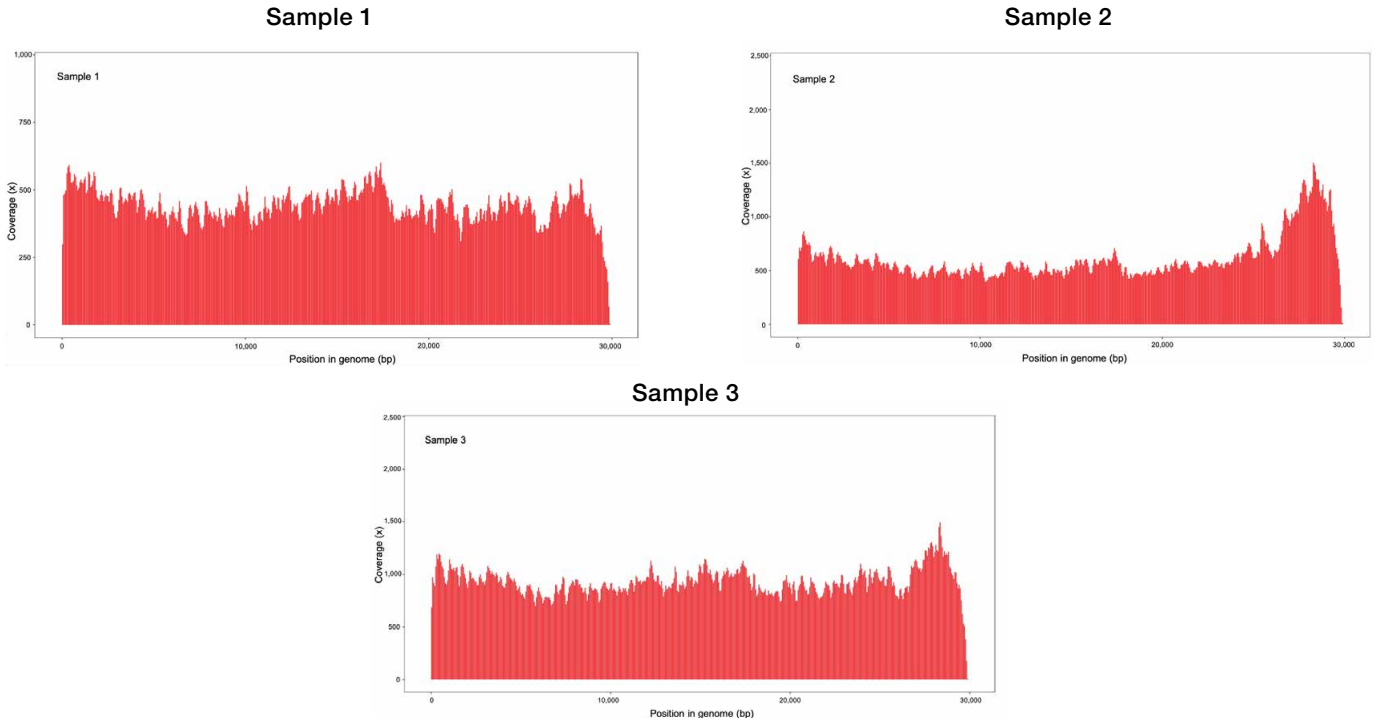
Final concentrations of sequencing libraries were determined using the **Invitrogen™ Collibri™ Library Quantification Kit**. Libraries were sequenced on the Illumina™ MiSeq™ System by paired-end sequencing of 2 x 150 bp, including a 10% v/v PhiX sequencing control. Alignment was calculated by QualiMap BamQC.

Results

Strong coverage and sensitive variant detection of three different BAL SARS-CoV-2 samples were accomplished with low depth of sequencing. Mean coverage in excess of 400x was demonstrated with less than 250,000 reads (Figure 2). The Collibri ES DNA Library Prep Kit for Illumina Systems was adequately sensitive for detecting single-nucleotide polymorphisms (SNPs) at low coverage. Sample quality corresponded to percentage of aligned reads, with all samples generating sufficient coverage to detect SNPs.

Conclusions

- NGS of SARS-CoV-2 can enable faster and more sensitive analysis of the viral genome and mutations. This can lead to a better understanding of viral evolution, transmission, and host–pathogen interactions.
- The optimized method described in this application note for the Collibri ES DNA Library Prep Kit enables sensitive detection of variants and even genome coverage for SARS-CoV-2 samples.
- Enrichment is optional, depending on the research goal. It may be required for genomic evolution studies but is not recommended for host–pathogen studies.



	Sample 1	Sample 2	Sample 3
Mean coverage	440.7x	612.1x	932.6x
Aligned reads (count)	96,059	134,431	209,029
Aligned reads (%)	99.5	99.6	99.7
SNP calls (count)	10	12	10

Figure 2. Strong coverage and sensitive detection of variants. Low coverage generates sufficient reads to identify SNPs against the reference **NCBI ASM985889v3 genome**. Coverage profiles from three research samples obtained from donors who tested positive for SARS-CoV-2 demonstrate coverage of the entire genomes from less than 250,000 reads per sample. The sensitivity of variant detection is suitable for strain identification of individual samples.

invitrogen

Key components

Step	Kit and link to user guide	Quantity	Cat. No.
Purification of total RNA*	MagMAX Viral/Pathogen Nucleic Acid Isolation Kit	Up to 2,000 preps	A48310
		Up to 200 preps	A42352
cDNA synthesis	SuperScript IV VILO Master Mix	50 reactions	11756050
		500 reactions	11756500
	Second Strand cDNA Synthesis Kit	50 reactions	A48570
		200 reactions	A48571
Library generation	Collibri ES DNA Library Prep Kit for Illumina Systems, with Combinatorial Dual (CD) Indexes	24 preps	A38605024
		96 preps	A38607096
	Collibri ES DNA Library Prep Kit for Illumina Systems, with Unique Dual (UD) Indexes	24 preps	A38606024 (Set A, 1-24)
			A43605024 (Set B, 25-48)
			A43606024 (Set C, 49-72)
			A43607024 (Set D, 73-96)
Library quantification**	Collibri Library Quantification Kit	100 reactions	A38524100
		500 reactions	A38524500
	Qubit dsDNA BR Assay Kit	100 assays	Q32850
		500 assays	Q32853

* Optional protocol modifications for reduced sample and reagent usage are described in the application note "**MagMAX Viral/Pathogen kit protocol changes to enable increased SARS-CoV-2 testing throughput**".

** Library quantification can be performed using the Collibri Library Quantification Kit (qPCR assay) or the Qubit dsDNA BR Assay Kit (fluorometric assay).

References

1. World Health Organization. Coronavirus. **who.int/health-topics/coronavirus#tab=tab_1**
2. Bedford T et al. (2020) Cryptic transmission of SARS-CoV-2 in Washington State. Preprint. *medRxiv* 2020.04.02.20051417.

Find out more at **thermofisher.com/sarscov2-collibri**

ThermoFisher SCIENTIFIC