illumina®

Detection and characterization of respiratory viruses, including SARS-CoV-2, using Illumina RNA Prep with Enrichment

A comprehensive and rapid target enrichment sequencing workflow for highly sensitive detection and characterization of common respiratory viruses, including COVID-19 strains.

Introduction

Viral infections are a major global health concern, with new infectious diseases continuing to emerge. The 2019 outbreak of novel coronavirus (SARS-CoV-2) that began in Wuhan, China and quickly spread to multiple countries is a particularly concerning example. Coronaviruses (CoV) are a large family of viruses that can infect humans, causing respiratory illnesses ranging from the common cold to more severe diseases, such as Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV). SARS-CoV-2 is a new strain not previously identified in humans. With millions of confirmed cases around the world and a death toll that has surpassed the SARS epidemic of 2003, the World Health Organization (WHO) has declared the disease associated with SARS-CoV-2 (COVID-19) a public health emergency of international concern, highlighting the need for rapid, accurate viral detection. Furthermore, research has shown higher rates of co-infection with SARS-CoV-2 and other viral pathogens than previously reported, which may impact disease management.[1]

Next-generation sequencing (NGS) provides an effective, novel way to screen samples and detect viruses without previous knowledge of the infectious agent.[2] Target enrichment is a resequencing method that captures genomic regions of interest by hybridization to target-specific biotinylated probes. Target enrichment through hybrid–capture methods allows for highly sensitive detection, without requiring the high read depth needed for shotgun metagenomics sequencing. Additionally, the target enrichment NGS workflow allows for near-complete sequence data of targets and opens up applications such as variant analysis for viral evolution or viral surveillance.[3] Compared to other targeted resequencing methods, such as amplicon sequencing, enrichment through hybrid capture allows for dramatically larger probe panels with more comprehensive profiling of the target regions. Additionally, the oligo probes used for hybrid–capture protocols remain effective, even within highly mutagenic regions, allowing targeting of rapidly evolving viruses, such as RNA viruses. In contrast to other methods used for viral detection, such as PCR or quantitative PCR (qPCR), NGS offers numerous advantages, including:

- Understand viral evolution
- Determine source of infection and route of transmission
- Identify and characterize co-infections and role of complex disease
- Screen targets for possible therapeutics

Integrated, comprehensive RNA enrichment workflow for respiratory virus detection

In response to the rapidly developing COVID-19 pandemic, Illumina previously provided guidance for studying COVID-19 using a modified protocol that included a third-party cDNA synthesis kit and Nextera™ Flex for Enrichment library preparation and hybrid capture.[4] However, with the launch of Illumina RNA Prep with Enrichment and additional DRAGEN™ analysis pipelines, Illumina offers an integrated, comprehensive, and fit-for-purpose solution for detection and characterization of respiratory viruses, including COVID-19 strains.

This application note highlights a streamlined workflow for detecting and analyzing COVID-19 using Illumina RNA Prep with Enrichment combined with the Illumina Respiratory Virus Oligos Panel v2, proven Illumina sequencing, and simplified data analysis (Figure 1).

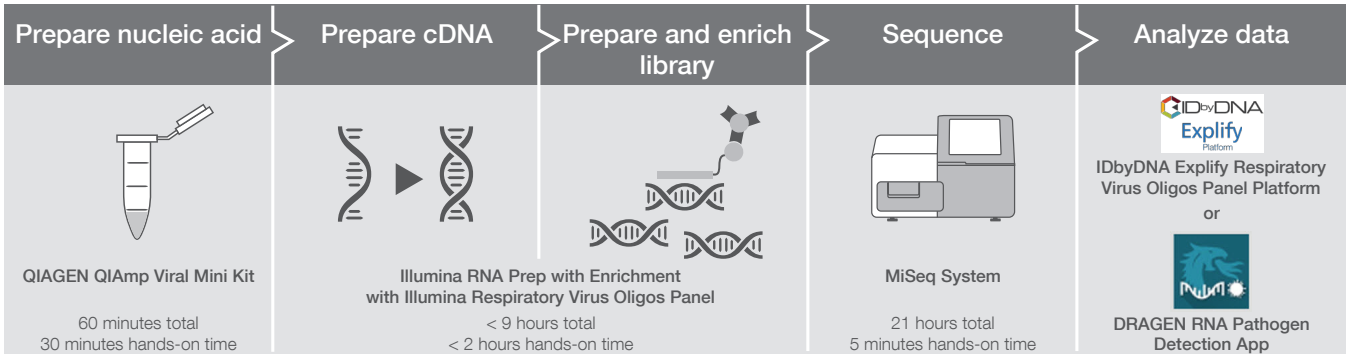

Figure 1: Enrichment workflow for respiratory virus detection—The streamlined NGS workflow for respiratory virus detection integrates sample preparation, library preparation, target enrichment, sequencing, and data analysis.

Table 1: Sequencing systems recommended for use with Illumina RNA Prep with Enrichment and the Respiratory Virus Oligos Panel

System	No. of cycles	Reads (M)	Run time (hours)	Samples per run (1M reads per sample)
NextSeq 550 System (high output)	150	400	18	400 max[a]
NextSeq 550 System (mid output)	150	130	15	130
MiSeq System (v3 chemistry)	150	25	21	25
MiniSeq System (high output)	150	25	17	25

a. Indicates that the number of samples per run exceeds the available number of indexes.

Illumina RNA Prep with Enrichment

Illumina RNA Prep with Enrichment uses On-Bead Tagmentation followed by a single hybridization step to provide a rapid workflow for generation of enriched RNA libraries. Illumina RNA Prep with Enrichment features multiple innovations that provide:

- Rapid, automation-compatible workflow that can be completed in a single day with minimal hands-on time
- Flexible support for input amounts of 10-100 ng RNA
- Scalable throughput that supports multiplexing of up to 384 samples in a single run with unique dual indexes

Proven Illumina NGS

Prepared libraries can be sequenced on any Illumina instrument. The benchtop MiniSeq™, MiSeq™, and NextSeq™ 550 Systems are well suited due to the low read requirements for the panel (Table 1).

DRAGEN RNA Pathogen Detection pipeline

The DRAGEN RNA Pathogen Detection pipeline enables streamlined detection of viral pathogens, including SARS-CoV-2, using coverage- and k-mer-based approaches (Figure 2). The new pipeline uses existing functionality of the DRAGEN RNA-Seq aligner and RNA-specific analysis components for gene expression quantification and gene fusion detection. The DRAGEN RNA Pathogen Detection is preconfigured to analyze and detect respiratory viruses. The pipeline supports customizable options for reference genomes, including:

- Custom reference that combines human hg38 with 168 viral sequences from the Seattle Flu Study, including SARS-CoV-2
- Custom reference based on the Illumina Respiratory Virus Oligos Panel for enhanced analysis
- Customer-generated references based on other panels or databases added through the input form

After alignment is complete, quality control (QC) removes duplicates and low-quality reads. Coverage plots are then created to detect SARS-CoV-2 and other viral strains. Additional features are included to support variant calling and creation of consensus FASTA files for upload to public databases, such as the Global Initiative on Sharing All Influenza Data (GISAID). The DRAGEN RNA Pathogen Detection pipeline is available in BaseSpace™ Sequence Hub and the DRAGEN Server.

Methods

This workflow is intended to enrich viral targets from total nucleic acid extraction. Reverse transcription of extracted RNA and library preparation is performed using Illumina RNA Prep with Enrichment, followed by enrichment with the Illumina Respiratory Virus Oligos Panel. After sequencing on an Illumina NGS system, data analysis proceeds using the DRAGEN Pathogen Detection pipeline and IDbyDNA Explify Platform.

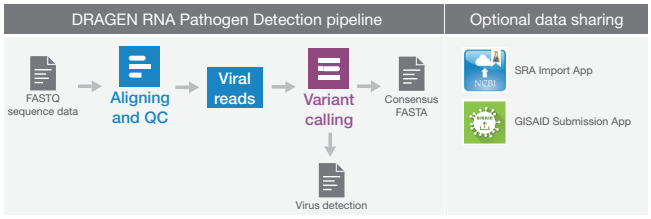

Figure 2: Analysis workflow for the DRAGEN RNA Pathogen Detection pipeline—Sequencing run data are input for alignment and QC, followed by detection of SARS-CoV-2 and other viral strains and variant calling. Consensus FASTA files can be output for upload to public databases.

Sample preparation

To demonstrate the performance of Illumina RNA Prep with Enrichment for detection of COVID-19, various strains of deactivated coronavirus sample were used in this study, including: CoV strain OC43 (QC Sets and Panels: Helix Elite; Microbiologics, Catalog no. 8217), Helix Elite Inactivated Standard (Microbiologics, Catalog no. HE0044N), AccuPlex SARS-CoV-2 Reference Material Kit (Sercare, Catalog no. 0505-0126), and Twist Synthetic SARS-CoV-2 RNA Control 1 (Twist, Catalog no. 102019). Samples* were extracted using the QIAGEN QIAmp Viral Mini Kit (QIAGEN, Catalog no. 52904) in a BSL2 laboratory environment (Table 1). Additionally, human saliva samples were collected and extracted to act as a human RNA background for the SARS-CoV-2 viral control RNA. 10 ng of extracted RNA was reversed transcribed into cDNA using reagents included with Illumina RNA Prep with Enrichment. Viral samples were run as extracted material or spiked into 95% Universal Human Reference (UHR) background RNA (Agilent Technologies, Catalog no. 74000) or human saliva RNA and reverse transcribed into cDNA to mimic sequencing results from patient samples (Table 1).

Additionally, a respiratory viral pool containing three RNA viruses (Table 2) was used to demonstrate detection of multiple RNA pathogens in one reactrion. The Microbiologics viral pool sample was tested with both a reaction volume of RNA (8.5 µl) and 1 µl of this sample spiked into 7.5 µl of UHR (10 ng UHR input). The viral pool sample was also spiked into 95% UHR background RNA (Agilent Technologies, Catalog no. 74000) and reverse transcribed into cDNA to mimic sequencing results from patient samples with multiple pathogens present (Table 1).

* Extraction was performed on a viral culture sample, not pure viral RNA.

Table 1: Composition of viral samples for analysis

Sample	Composition[a]	Reference genome
CoVOC43	10 ng CoV OC43 RNA	AY391777.1 Human coronavirus OC43
CoVOC43_95UHR	0.5 ng CoV OC43 RNA and 9.5 ng UHR RNA	AY391777.1 Human coronavirus OC43
ViralPool	8.5 µl of viral pool	See Table 2
ViralPool_95UHR	1 µl viral pool in 10 ng UHR background	See Table 2
SeraCare AccuPlex SARS-CoV-2 Reference Material Kit	7.5 µl sample in 10 ng UHR background	NC_045512
Twist Synthetic SARS-CoV-2 RNA Control 1	Titration of viral copies into 10 ng human saliva RNA background	MT007544.1

a. The recommended minimum RNA input for reverse transcription is 10 ng. For best results, reverse transcription should be performed on freshly extracted nucleic acid samples.

Table 2: Composition of Microbiologics viral pool

Virus
Influenza A virus (H1N1)
Influenza B virus, strain Hong Kong/5/72
Respiratory syncytial virus (RSV) A, strain long
Helix Elite Inactivated Standard (Catalog no. HE0044N)

Library preparation

Sequencing-ready libraries were prepared using cDNA from the CoV samples, the viral pool sample, and spike-in samples with Illumina RNA Prep with Enrichment (Illumina, Catalog no. 20040536) and IDT for Illumina DNA/RNA UD Indexes (Illumina, Catalog no. 20027213). Total RNA input recommended for tagmentation is 10-100 ng.

After amplification, samples were enriched as single-plex reactions using the Respiratory Virus Oligos Panel v2 (Illumina, Catalog no. 20044312), which features ~7800 probes designed to detect respiratory viruses, recent flu strains, and SARS-CoV-2, as well as human probes to act as positive controls in every reaction (Table 3). After enrichment, the libraries were prepped for sequencing.

Sequencing

Prepared libraries were denatured and diluted to a final loading concentration of 10 pM, according to the MiSeq System Denature and Dilute Libraries Guide (Document no. 15039740 v10) and sequenced on the MiSeq System at 2 × 75 bp read length using MiSeq v3 reagents. Aliquots of all libraries were also sequenced prior to probe hybridization to determine the fold change of the enrichment reaction (optional step to demonstrate the successful enrichment of the viral targets, not part of the standard workflow).

Virus titer, RNA quality, and the number of reads per sample impact the number of virus-specific reads obtained and coverage of the viral genome. As a general guideline, the read recommendation for this workflow is 1M reads per sample, but these numbers can be variable and this is only a recommended starting point.

Table 3: Viruses targeted by the Respiratory Virus Oligos Panel

Human coronavirus 229E	Human parainfluenza virus 1
Human coronavirus NL63	Human parainfluenza virus 2
Human coronavirus OC43	Human parainfluenza virus 3
Human coronavirus HKU1	Human parainfluenza virus 4a
SARS-CoV-2	Human metapneumovirus (CAN97-83)
Human adenovirus B1	Respiratory syncytial virus (type A)
Human adenovirus C2	Human Respiratory syncytial virus 9320 (type B)
Human adenovirus E4	Influenza A virus (A/Puerto Rico/8/1934(H1N1))
Human bocavirus 1 (Primate bocaparvovirus 1 isolate st2)	Influenza A virus (A/Korea/426/1968(H2N2))
Human bocavirus 2c PK isolate PK-5510	Influenza A virus (A/New York/392/2004(H3N2))
Human bocavirus 3	Influenza A virus (A/goose/Guangdong/1/1996(H5N1))
Human bocavirus 4 NI strain HBoV4-NI-385	Influenza A virus (A/Zhejiang/DTID-ZJU01/2013(H7N9))
KI polyomavirus Stockholm 60	Influenza A virus (A/Hong Kong/1073/99(H9N2))
WU Polyomavirus	Influenza A virus (A/Texas/50/2012(H3N2))
Human parechovirus type 1 PicoBank/HPeV1/a	Influenza A virus (A/Michigan/45/2015(H1N1))
Human parechovirus 6	Influenza B virus (B/Lee/1940)
Human rhinovirus A89	Influenza B virus (B/Wisconsin/01/2010)
Human rhinovirus C (strain 024)	Influenza B virus (B/Brisbane/60/2008)
Human rhinovirus B14	Influenza B virus (B/Colorado/06/2017)
Human enterovirus C104 strain: AK11	Influenza B virus (B/Washington/02/2019)
Human enterovirus C109 isolate NICA08-4327	Human control genes

Data analysis

FASTQ sequencing data files were input to the DRAGEN RNA Pathogen Detection pipeline and the IDbyDNA Explify Respiratory Virus Oligos Panel Platform for analysis and viral detection. Data analysis is based on k-mers and alignment steps, including protein-encoding transcript level detection of viruses, which increases the ability to identify novel and highly divergent viruses. These platforms can be accessed in BaseSpace Sequence Hub.

Results

To demonstrate the exceptional performance of the RNA enrichment workflow for respiratory virus detection, various COVID-19 control samples were evaluated.

Evaluation of COVID-19 detection with the DRAGEN RNA Pathogen Detection pipeline

Improved detection and genome coverage of SARS-CoV-2 in viral titration experiment

A titration experiment was conducted in which Twist Synthetic SARS-CoV-2 RNA Control 1 was spiked into UHR RNA in diminishing amounts of virus, from 1M down to 0 viral copies. 10 ng total RNA was input for reverse transcription, libraries were prepared and enriched from the resulting cDNA, and sequencing was performed on the MiSeq System, normalized to 1M reads. The same samples were also run after cDNA conversion and library preparation with Nextera Flex for Enrichment and the Respiratory Virus Oligos Panel v1 for performance comparison. Analysis with the DRAGEN RNA Pathogen Detection pipeline successfully identified SARS-CoV-2, even at low viral copy numbers (Figure 3).

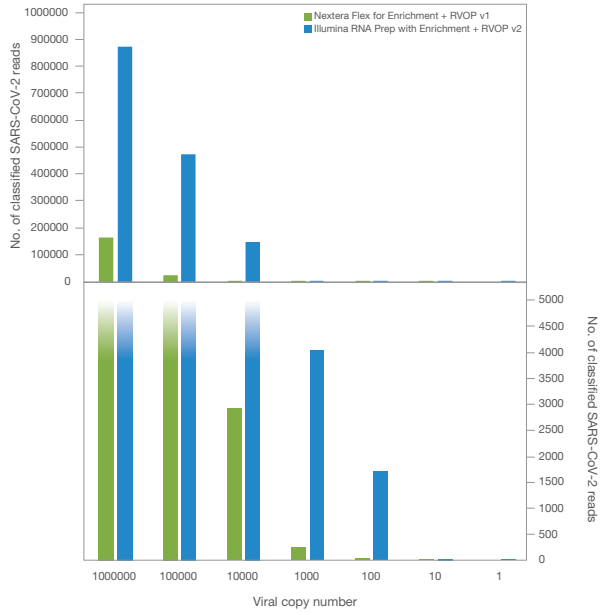

Figure 3: Improved SARS-CoV-2 detection using Illumina RNA Prep with Enrichment—In a titration experiment with diminishing amounts (1M down to 0 copies) of virus spiked into UHR RNA, Illumina RNA Prep with Enrichment resulted in improved detection of SARS-CoV-2, as compared to Nextera Flex for Enrichment. The top plot shows the full range of data for all viral amounts; the bottom plot focuses in on the lower viral amounts.

Nearly full SARS-CoV-2 genome coverage was achieved at 1000 copies, with higher depth achieved at 10,000 copies, and minimal coverage achieved with as low as 1-10 copies (Figure 4). All testing was performed at the minimum input of 10 ng of RNA per sample, but higher input would increase library diversity and coverage, if higher input is available (data not shown).

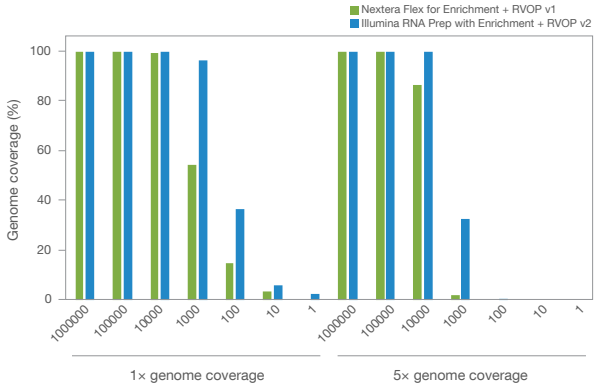

Figure 4: Improved SARS-CoV-2 genome coverage using Illumina RNA Prep with Enrichment—Nearly full SARS-CoV-2 genome coverage was achieved at 1000 copies, with higher depth achieved at 10,000 copies, and minimal coverage achieved with as low as 1-10 copies.

Highly sensitive SARS-CoV-2 detection in granular viral titration experiment

In order to better understand the sensitivity of the assay, an additional titration with between 2 and 1000 copies was performed with an increased number of data points showing detection of viral reads as low as 2 copies (Figure 5).

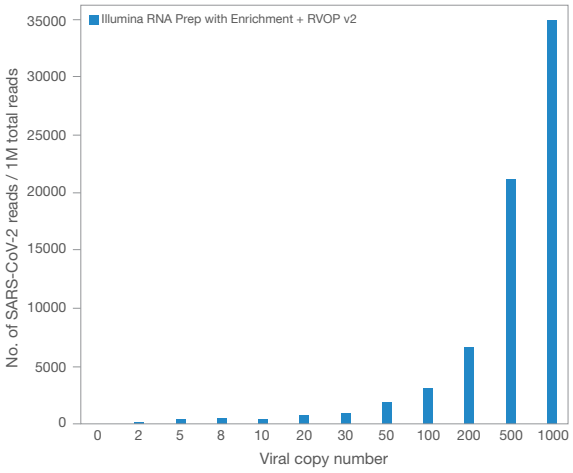

Figure 5: Highly sensitive SARS-CoV-2 detection using Illumina RNA Prep with Enrichment—In a titration experiment with 1000 down to two copies of virus spiked into human saliva RNA, Illumina RNA Prep with Enrichment resulted in detection of viral reads from as low as two copies of SARS-CoV-2.

Evaluation of viral target detection with additional commercial samples

The percent aligned reads from commercially available viral control samples pre- and post-enrichment were compared and used to calculate the fold change of enrichment, as a measure of the success of the target capture and hybridization reaction. Results showed significantly more aligned reads for viral strains in the Helix Elite Inactivated Standard (Microbiologics, Catalog no. HE0044N), including respiratory syncytial virus A, influenza A, influenza B (Figure 6) and OC43 (Figure 7). Similar increases were seen for SARS-CoV-2 with the AccuPlex SARS-CoV-2 Reference Material Kit (SeraCare, Catalog no. 0505-0126) after enrichment (Figure 7). Fold enrichment calculations showed significant increases for the commercially available control samples analyzed (Table 4).

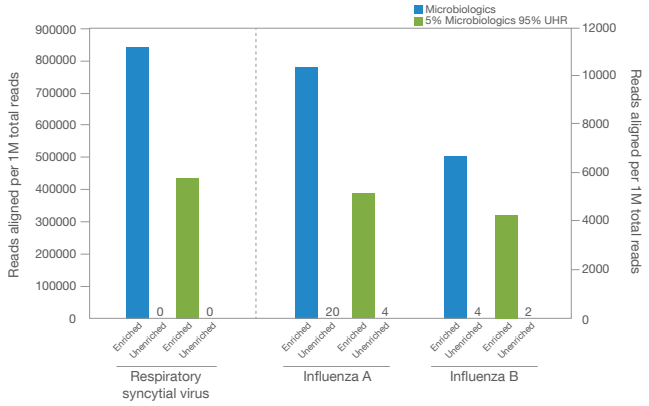

Figure 6: Improved respiratory viral detection using Illumina RNA Prep with Enrichment—Illumina RNA Prep with Enrichment resulted in successful detection of viral strains in the Helix Elite Inactivated Standard from Microbiologics.

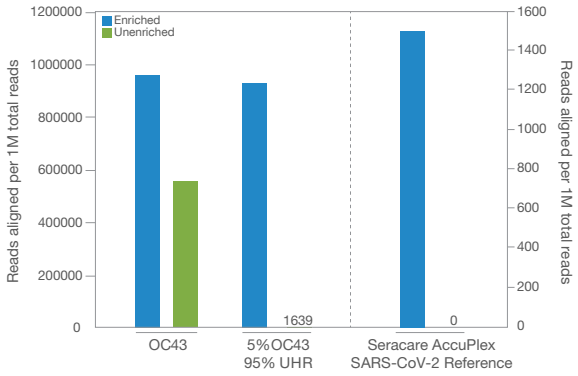

Figure 7: Improved SARS-CoV-2 detection using Illumina RNA Prep with Enrichment—Illumina RNA Prep with Enrichment resulted in successful detection of the CoVOC43 strain in the Helix Elite Inactivated Standard from Microbiologics (left) and SARS-CoV-2 in the AccuPlex SARS-CoV-2 Reference Material Kit from Seracare (right).

Table 4: Fold enrichment with Respiratory Virus Oligos Panel

Commercial control	Fold enrichment	
	Influenza A	Influenza B
Microbiologics	517×	1117×
Microbiologics_95UHR	1277×	2123×
	Respiratory syncytial virus[a]	
Microbiologics	Detected	
Microbiologics_95UHR	Detected	
	OC43 human coronavirus	
OC43	2×	
OC43_95UHR	580×	

a. RSV was not detected in unenriched libraries, so fold enrichment cannot be calculated. Mean coverage metrics have been normalized to 1M reads per sample.

Analysis with IDbyDNA Explify

The IDbyDNA Explify Respiratory Virus Oligos Panel Platform identified viral strains in the Helix Elite Inactivated Standard from Microbiologics (Figure 8A-C) and CoVOC43 (Figure 8C-E).

The IDbyDNA Explify Respiratory Virus Oligos Panel Platform also confirmed that Illumina RNA Prep with Enrichment and the Respiratory Virus Oligos Panel v2 resulted in improved detection and genome coverage of SARS-CoV-2, as compared to Nextera Flex for Enrichment and the Respiratory Virus Oligos Panel v1 (Figure 9).

Evaluation of enrichment plexity on performance

To evaluate the effect of plexity, the number of pre-enriched libraries that are pooled together in a single enrichment reaction, on SARS-CoV-2 detection, libraries with varying numbers of viral copies (100 and 1M copies) were prepared with Illumina RNA Prep with Enrichment and enriched with the Respiratory Virus Oligos Panel v2 either at 1-plex or 3-plex. Three-plex libraries were pooled either in an unbalanced or balanced fashion. SARS-CoV-2 was detected in all samples at both single plex and a balanced 3-plex (Figure 10, Table 5).

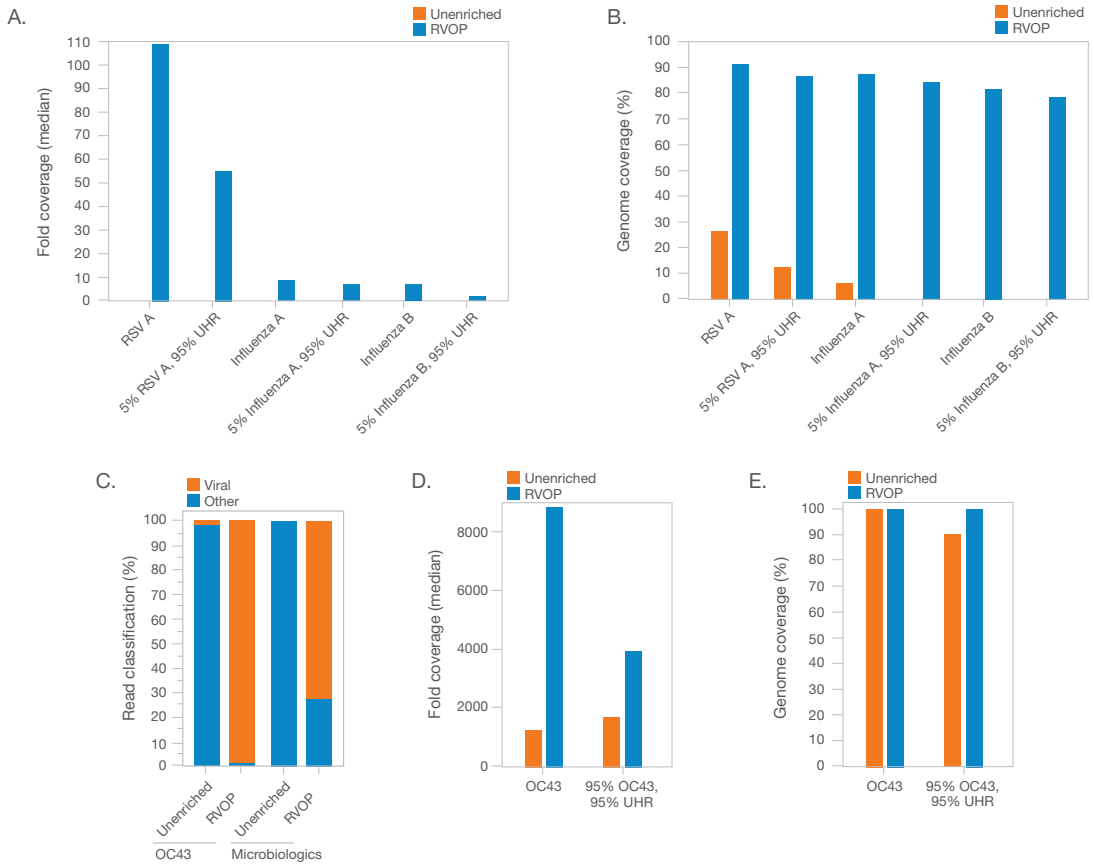

Figure 8: Identification of respiratory virus strains with IDbyDNA Explify—The IDbyDNA Explify Respiratory Virus Oligos Platform was used to successfully identify respiratory virus strains in the Helix Elite Inactivated Standard from Microbiologics, as measured by A) fold coverage B) genome coverage, and C) read classification. The platform was also used to successfully identify OC43, as measured by C) read classification, D) fold coverage, and E) genome coverage.

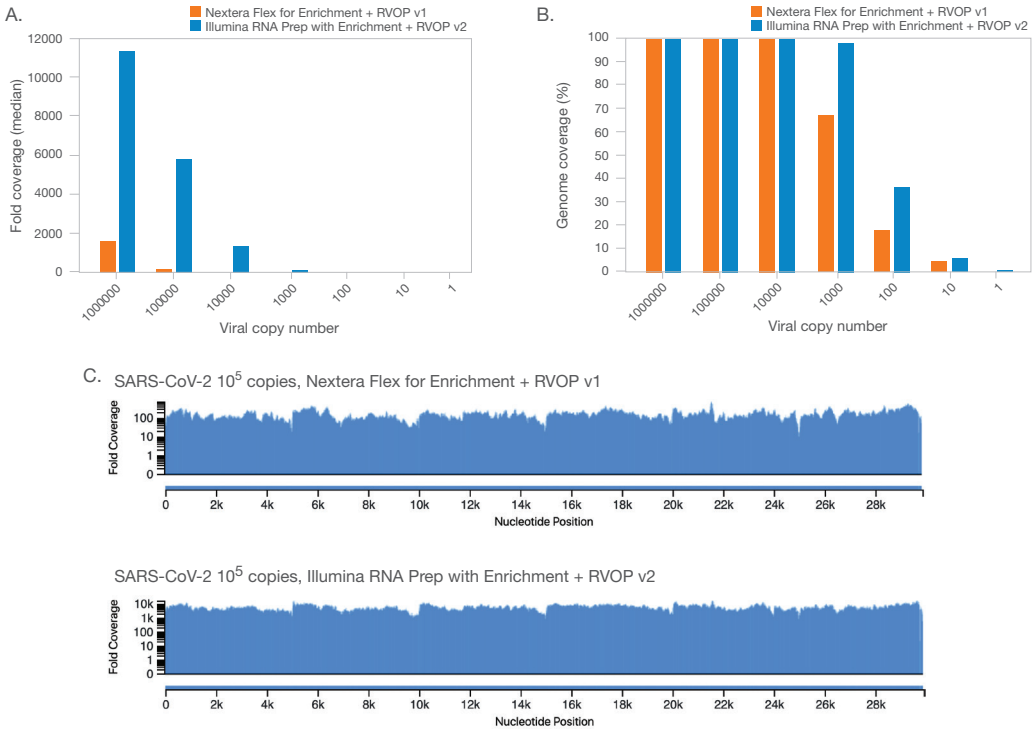

Figure 9: Identification of SARS-CoV-2 with IDbyDNA Explify—Analysis of the titration experiment with diminishing amounts (1M down to 0 copies) of virus spiked into UHR RNA with IDbyDNA Explify Respiratory Virus Oligos Platform successfully identified SARS-CoV-2 with increased A) fold coverage and B, C) genome coverage.

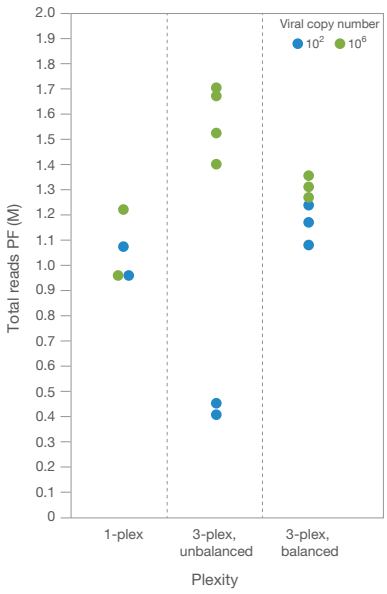

Figure 10: SARS-CoV-2 detection using Illumina RNA Prep with Enrichment at different plexities—Illumina RNA Prep with Enrichment libraries with 10^2 and 10^6 viral copy numbers were enriched with the Respiratory Virus Oligos Panel v2 at 1- or 3-plex reactions. SARS-CoV-2 was detected in all samples at both 1-plex and 3-plex, balanced. In each scenario, 1.1M reads were targeted per sample. The percent viral reads from samples containing 10^2 viral copies ranged from 0.1% to 0.3% while the samples containing 10^6 viral copies ranged from 66% to 89%

Table 5: SARS-CoV-2 total reads PF at different plexities

Plexity	Viral copy number	Total R1 reads PF
1-plex	1M[e]	1,222,000
	1M[e]	965,080
	100	952,120
	100	1,074,620
3-plex, unbalanced	1M[a]	1,704,660
	1M[a]	1,525,820
	1M[b]	1,671,340
	1M[b]	1,401,720
	100[a]	454,140
	100[b]	409,740
3-plex, balanced	1M[c]	1,309,310
	1M[c]	1,354,190
	1M[c]	1,257,570
	100[d]	1,238,530
	100[d]	1,078,480
	100[d]	1,169,280

a. Libraries were pooled together for 3-plex, unbalanced reaction.
b. Libraries were pooled together for 3-plex, unbalanced reaction.
c. Libraries were pooled together for 3-plex, balanced reaction.
d. Libraries were pooled together for 3-plex, balanced reaction.

Summary

The identification and characterization of emerging viruses is central to improving public health. In these situations, NGS is a powerful method for broad-range detection to identify known and emerging viruses. Using Illumina RNA Prep with Enrichment with the Respiratory Virus Oligos Panel enables researchers to obtain genomic data that can confirm the presence of SARS-CoV-2 and advance studies such as genotyping and variant analysis. The agnostic design allows for widespread identification of pathogenic viruses across all sample types of interest and the use of unique dual indexes reduces the risk of any indexing crossover from multiplexing samples. This easy-to-follow workflow, enables detection and characterization of pathogen outbreaks, such as COVID-19.

Learn more

Learn more about viral sequencing methods at www.illumina.com/areas-of-interest/microbiology/infectious-disease-surveillance.html

Access publicly available data sets from experiments described in this application note in BaseSpace Sequence Hub by searching for "RVOPv2".

References

1. Kim D, Quinn J, Pinksy B, Shah NH, Brown I. Rates of co-infection between SARS-Cov-2 and other respiratory pathogens. *JAMA*. 2020; 323(20):2085-2086.
2. Bulcha B. Review on viral metagenomics and its future perspective in zoonotic and arboviral disease surveillance. *J Biol Agric Healthcare*. 2017;7(21):35–41.
3. Gaudin M and Desnues C. Hybrid Capture-Based Next Generation Sequencing and Its Application to Human Infectious Diseases. *Front Microbiol*. 2018;9:2924.
4. Illumina (2020) Enrichment workflow for detecting coronavirus using Illumina NGS systems. Accessed May 7, 2020.

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

illumina®