Celemics Comprehensive Respiratory Virus Panel (CRVP)

Detection panel for the accurate sequencing and identification of prominent whole viral genomes, including coinfection possibilities such as coronavirus and influenza.

Introduction

Viral respiratory diseases have become an increasingly dire global health concern, and epidemics such as MERS, SARS, and Ebola have shown the need for rapid identification of viral strains in order to expedite treatment. The recent COVID-19 pandemic has only magnified the need for quick and accurate identificationfor the understanding of respiratory viruses. Additionally, in order to prevent any possible confusion in the identification of specific respiratory virus strains, there has also been an increase in the need of observing whole virus genomes, veering away from PCR-based assays, where detection was based on the targeting of conserved regions of pathogen genomes1.

The Celemics Comprehensive Respiratory Virus Panel (CRVP) was developed as a targeted respiratory virus panel that can not only detect and identify specific virus strains, but can also sequence whole viral genomes. The targeted nature of the panel allows for the highly efficient production of sequencing data and the subsequent procurement of credible results. Targeting the entire viral genome enables the generation of consensus sequences and the identification of novel viral mutations.

Virus Strains

Major respiratory viral strains were chosen based on their known prevalence and proclivity to cause respiratory diseases in humans (Table 1).

Different subtypes of viruses were also accounted for where their identification and treatment may require an alternative approach. Complete viral genome sequences were retrieved from the NCBI RefSeq database.

Table 1. Virus strains used in CRVP. Viral strains chosen based on their prevalence in humans, with genomes acquired from NCBI RefSeq database.

Virus	Strain	Virus	Strain
Human Adenovirus	Human Adenovirus Type 1 (HAdV-C1)	Human Enterovirus	EV-C104
	Human Adenovirus Type 2 (HAdV-C2)		EV-C105
	Human Adenovirus Type 3 (HAdV-B3)		EV-C109
	Human Adenovirus Type 4 (HAdV-E4)		EV-C117
	Human Adenovirus Type 5 (HAdV-C5)		EV-C118
	Human Adenovirus 7 (HAdV-B7)		CV-A21
	Human Adenovirus 14 (HAdV-B14)		EV-D68
	Human Adenovirus 21 (HAdV-B21)	Influenza A	Influenza A Virus (Flu A)
Bocavirus 1/2/3/4 (HBoV)	Human Bocavirus 1		Influenza A-H1 Virus (Flu A-H1)
	Human Bocavirus 2		Influenza A-H3 Virus (Flu A-H3)
	Human Bocavirus 3	Influenza B	Influenza B Virus (Flu B)
	Human Bocavirus 4	Parainfluenza Virus	Parainfluenza 1 (PIV 1)
Human Rhinovirus (A/B/C)	Human Rhinovirus A		Parainfluenza 2 (PIV 2)
	Human Rhinovirus B		Parainfluenza 3 (PIV 3)
	Human Rhinovirus C		Parainfluenza 4 (PIV 4) A
Coronavirus	Coronavirus HKU1		Parainfluenza 4 (PIV 4) B
	Coronavirus NL63	Respiratory Syncytial Virus	Respiratory Syncytial Virus A (RSV A)
	Coronavirus 229E		Respiratory Syncytial Virus B (RSV B)
	Coronavirus OC43		Human Metapneumovirus
	SARS-CoV-2		

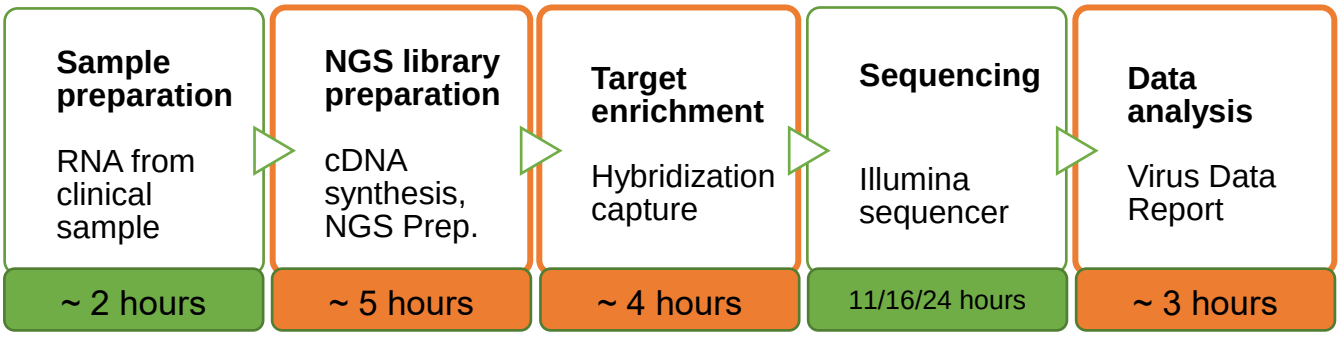

Figure 1. Overall Workflow for Celemics CRVP. Portions of the procedure covered through the Complete Kit are hlighted in orange. Data analysis only includes an interactive report based on sequencing results.

Methods

Introduction

Universal Human Reference was used as a negative control, and synthetic RNA virus samples acquired from Twist Biosciences were used for analytical validation. All specimens were obtained in RNA form from clinical samples that tested positive for COVID-19, as the Celemics kit used starts off sample preparation using RNA.

After initial QC was performed using an Agilent 4200TapeStation System, clinical samples with an RNA input amount of 25 ng or less had carrier RNA (UHR) added. Analytical validation samples were made by adding 100 ng UHR to synthetic RNA depending on copy number, then processed through cDNA synthesis, from which fragmented RNA was converted first to a cDNA/RNA hybrid (1st strand synthesis) then to complete cDNA (2nd strand synthesis).

NGS Library Preparation

Sequencing-ready libraries were created from the prepared cDNA using a Celemics Library Prep Kit. The cDNA was amplified after adapter ligation and end repair, then pre-pooled into groups of 8 samples, each with 200ng, for a total of 1600ng. For samples of lackluster quality, pooling was conducted by adding the maximum possible amount. For analytical validation samples, 200 ng of each sample was individually captured.

Target Enrichment

After samples were amplified and run through an additional Quality Check to assess data quality, all samples were processed through hybridization using the Celemics Target Enrichment Kit. The end quantity of each sample's sequencing data was 0.4 Gb.

Sequencing

Completed libraries were sequenced using the Illumina NextSeq 500 System and the Mid/High Output Kit v2.5 (300 cycles). Sequencing preparation was conducted according to the IlluminaNextSeq Running Protocol at 2 x 150bp read length,but can also be run at 2 x 75bp and 1 x 75bp as well, depending on the required experimental time.

Data Analysis

The Celemics Virus Verifier, which can identify and generate consensus sequences for viral genomes, was created to analyze FASTQ files generated from the sequencer. It allows for easy initiation of analysis through a Linux-based operating system. Once analysis is launched, the FASTQ file is filtered and trimmed. Human DNA sequences are removed from the filtered FASTQ for more accurate identification and expedited analysis. Using CRVP's viral genome, the Celemics Virus Verifier identifies the virus genome and quantity present in the givensample (Figure 2).

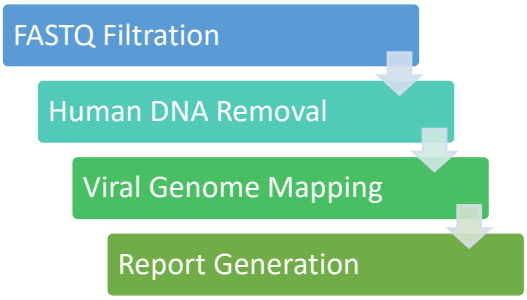

Figure 2. Virus Analysis Workflow. A brief overview of the post-sequencing procedures taken for bioinformatics analysis.

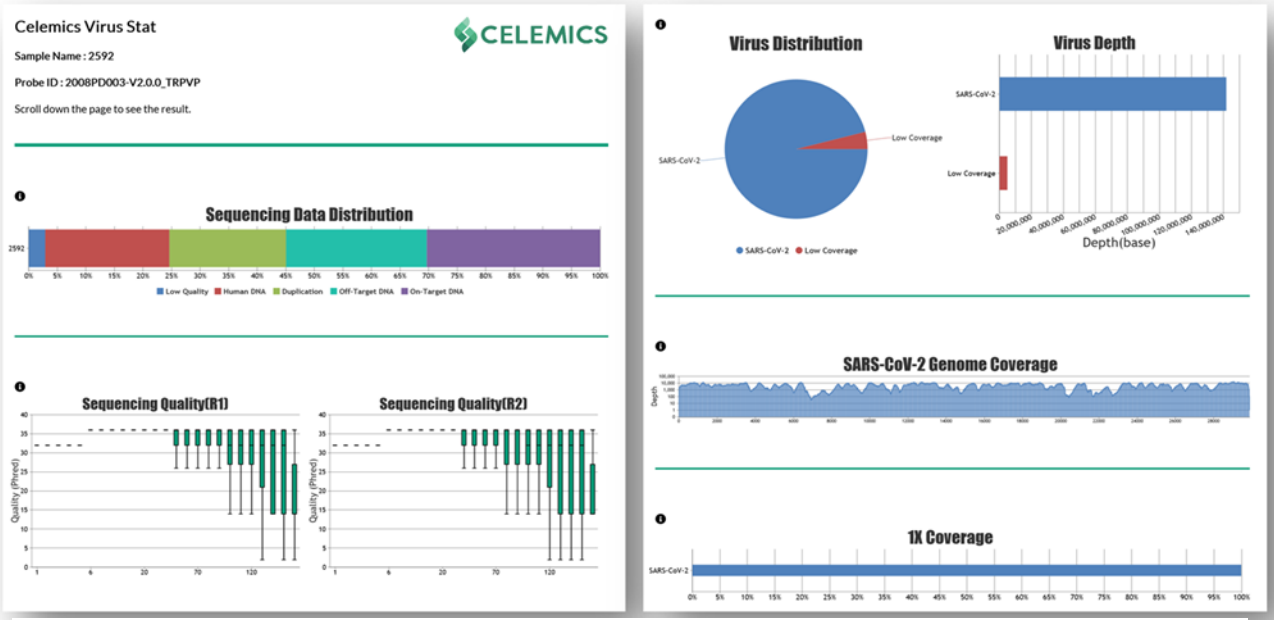

Figure 3. Example Celemics Virus Report. Data generated using a clinical sample set. Report generated as an interactable PDF for ease of access and inclusion of all relevant datasets.

Results

Ability to Identify Virus Among Multiple Viral Strains

39 in-silico viral FASTQ files containing reads from each virus strain were analyzed with the Celemics Virus Verifier, and results showed that 100% of reads were aligned to their respective counterparts and none were mismatched (Table 2). Further testing was performed by generating a single sample with all 39 viral reads combined together. The Celemics Virus Verifier was able to accurately identify each read to its origin without error (Figure 4).

Table 2. Identification of 39 in-silico virus samples

Virus Type	Mapped Correctly	Mapped Incorrectly	1X Coverage
Coronavirus_229E	100.00%	0.00%	100.00%
Coronavirus_HKU1	100.00%	0.00%	100.00%
Coronavirus_NL63	100.00%	0.00%	100.00%
Coronavirus_OC43	100.00%	0.00%	100.00%
Human_adenovirus_14_HAdV-B14	100.00%	0.00%	100.00%
Human_adenovirus_21_HAdV-B21	100.00%	0.00%	100.00%
Human_adenovirus_type_3_HAdV-B3	100.00%	0.00%	100.00%
Human_adenovirus_7_HAdV-B7	100.00%	0.00%	100.00%
Human_adenovirus_type_1_HAdV-C1	100.00%	0.00%	99.33%
Human_adenovirus_type_2_HAdV-C2	100.00%	0.00%	99.13%
Human_adenovirus_type_5_HAdV-C5	100.00%	0.00%	99.79%
Human_adenovirus_type_4_HAdV-E4	100.00%	0.00%	100.00%
Human_bocavirus_1	100.00%	0.00%	100.00%
H uman_bocavirus_2	100.00%	0.00%	100.00%
Human_bocavirus_3	100.00%	0.00%	100.00%
Human_bocavirus_4	100.00%	0.00%	100.00%
Human_Enterovirus_CV-A21	100.00%	0.00%	100.00%
Human_Enterovirus_EV-C104	100.00%	0.00%	100.00%
Human_Enterovirus_EV-C105	100.00%	0.00%	100.00%
Human_Enterovirus_EV-C109	100.00%	0.00%	100.00%
Human_Enterovirus_EV-C117	100.00%	0.00%	100.00%
Human_Enterovirus_EV-D68	100.00%	0.00%	100.00%
Human_Enterovirus_EV-C118	100.00%	0.00%	100.00%
Human_Metapneumovirus	100.00%	0.00%	100.00%
Human_Rhinovirus_A	100.00%	0.00%	100.00%
Human_Rhinovirus_B	100.00%	0.00%	100.00%
Human_Rhinovirus_C	100.00%	0.00%	100.00%
Influenza_virus_A_H7N9	100.00%	0.00%	100.00%
Influenza_virus_A-H1_H1N1	100.00%	0.00%	100.00%
Influenza_virus_A-H3_H3N2	100.00%	0.00%	100.00%
Influenza_virus_B	100.00%	0.00%	100.00%
Parainfluenza_virus_1	100.00%	0.00%	100.00%
Parainfluenza_virus_2	100.00%	0.00%	100.00%
Parainfluenza_virus_3	100.00%	0.00%	100.00%
Parainfluenza_virus_4A	100.00%	0.00%	100.00%
Parainfluenza_virus_4B	100.00%	0.00%	100.00%
Respiratory_syncytial_virus_A	100.00%	0.00%	100.00%
Respiratory_syncytial_virus_B	100.00%	0.00%	100.00%
SARS-CoV-2	100.00%	0.00%	100.00%

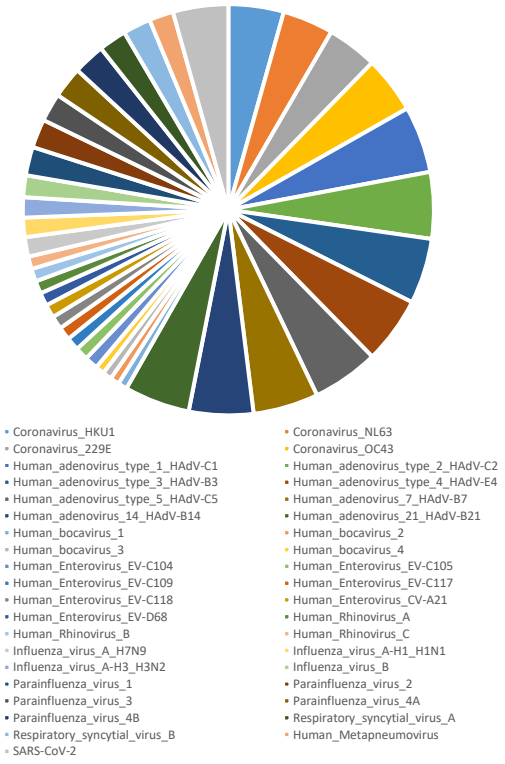

Figure 4. Virus Distribution of In-Silico 39 Virus Mixed Sample

Sensitivity of CRVP

To test CRVP's sensitivity, 6 different virus samples were used. Each virus sample was diluted into three different concentrations, and each condition was triplicated. Result show that on average, CRVP can detect and identify samples with viral genome concentrations at 100 copies. Samples with 1,000,000 copies have shown high depth and over 99% 1x coverage, which is enough to generate whole genome consensus sequences (Table 3). Note that poor sample quality can cause low depth and coverage. SARS-CoV-2 (COVID-19) clinical samples acquired from the upper respiratory tract (URT, 9 samples), NP/OP swab (17 samples), and unknown samples (10 samples) were also tested. The CRVP panel accurately identified SARS-CoV-2 from every clinical sample and was able to generate whole genome consensus sequences from most samples (Table 4).

Additional experimentation was performed to measure coverage levels in samples with multiple infections. 6 species of non-clinical reference virus species (1,000,000 copies each) were mixed with 100ng UHR and sequenced, and results indicated that all six species had a 1X coverage level of at least 99.9%, showing CRVP's ability for multiple co-infection detection (Figure 5).

Table 3. Virus samples tested with varying degreesof copy numbers

Virus		100 Copies	10,000 Copies	1,000,000 Copies
Human Enterovirus EV-D68	Reads	38	1,677	192,026
	Coverage (1X)	25.68%	99.06%	100.00%
Parainfluenza virus 4A	Reads	18	2,066	248,756
	Coverage (1X)	6.31%	74.22%	99.98%
Influenza virus A-H1 H1N1	Reads	37	3,481	212,185
	Coverage (1X)	11.84%	94.37%	100.00%
Influenza virus A-H3 H3N2	Reads	39	4,375	251,184
	Coverage (1X)	16.29%	97.22%	99.99%
Influenza virus B	Reads	7	417	37,410
	Coverage (1X)	4.16%	63.74%	99.64%
SARS-CoV-2	Reads	8	633	45,512
	Coverage (1X)	1.43%	39.01%	98.85%

Table 4. Identification of virus strains in clinical samples (low quality samples omitted of copy numbers)

Sample Number	Sample Type	Virus	Mapped Correctly	Mapped Incorrectly	Coverage (1X)
1	URT	SARS-CoV-2	99.94%	0.06%	98.74%
2	URT	SARS-CoV-2	99.98%	0.02%	99.19%
3	URT	SARS-CoV-2	99.74%	0.26%	99.82%
4	URT	SARS-CoV-2	100.00%	0.00%	99.95%
5	URT	SARS-CoV-2	99.64%	0.36%	99.53%
6	URT	SARS-CoV-2	100.00%	0.00%	99.94%
7	URT	SARS-CoV-2	100.00%	0.00%	99.88%
8	URT	SARS-CoV-2	100.00%	0.00%	99.93%
9	URT	SARS-CoV-2	98.81%	1.19%	99.42%
10	NP/OP swab	SARS-CoV-2	99.61%	0.39%	99.86%
11	NP/OP swab	SARS-CoV-2	99.95%	0.05%	99.65%
12	NP/OP swab	SARS-CoV-2	99.99%	0.01%	94.13%
13	NP/OP swab	SARS-CoV-2	99.75%	0.25%	99.72%
14	NP/OP swab	SARS-CoV-2	99.99%	0.01%	99.86%
15	NP/OP swab	SARS-CoV-2	99.90%	0.10%	99.92%
16	NP/OP swab	SARS-CoV-2	99.92%	0.08%	99.89%
17	NP/OP swab	SARS-CoV-2	99.97%	0.03%	99.87%
18	NP/OP swab	SARS-CoV-2	100.00%	0.00%	99.90%
19	NP/OP swab	SARS-CoV-2	100.00%	0.00%	99.89%
20	NP/OP swab	SARS-CoV-2	100.00%	0.00%	99.87%
21	NP/OP swab	SARS-CoV-2	99.99%	0.01%	99.87%
22	NP/OP swab	SARS-CoV-2	100.00%	0.00%	99.86%
23	NP/OP swab	SARS-CoV-2	100.00%	0.00%	99.90%
24	NP/OP swab	SARS-CoV-2	99.31%	0.69%	98.87%
25	NP/OP swab	SARS-CoV-2	99.99%	0.01%	99.53%
26	NP/OP swab	SARS-CoV-2	99.99%	0.01%	99.79%
27	Unidentified	SARS-CoV-2	99.77%	0.23%	99.91%
28	Unidentified	SARS-CoV-2	100.00%	0.00%	99.95%
29	Unidentified	SARS-CoV-2	100.00%	0.00%	99.95%
30	Unidentified	SARS-CoV-2	99.99%	0.01%	99.94%
31	Unidentified	SARS-CoV-2	99.84%	0.16%	96.82%
32	Unidentified	SARS-CoV-2	100.00%	0.00%	89.03%
33	Unidentified	SARS-CoV-2	100.00%	0.00%	99.91%
34	Unidentified	SARS-CoV-2	99.92%	0.08%	94.48%
35	Unidentified	SARS-CoV-2	99.97%	0.03%	99.91%
36	Unidentified	SARS-CoV-2	100.00%	0.00%	99.88%

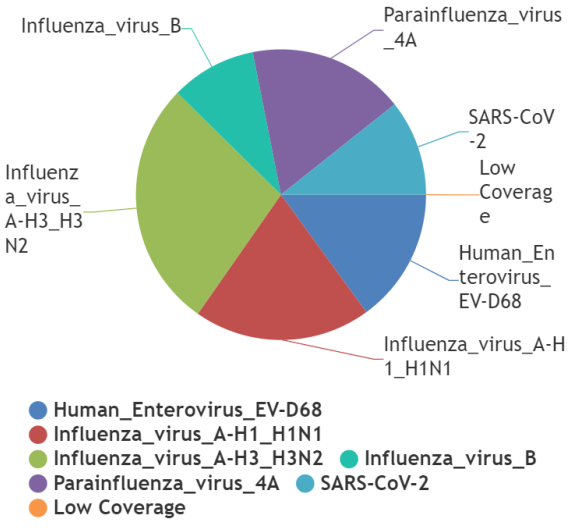

Figure 5. Virus Distribution of 6 reference virus samples

Coverage and Consensus Sequence of Viral Genome

Optimized capture efficiency of viral target probes enabled mean depths of over 99.89%, 99.86%, and 99.03% for 1X, 10X, 100X coverage of viral genome in a clinical sample using a 300 cycle (151 bp paired end) sequencing kit. To examine the effect of shorter reads on virus identification, 151 bp reads were cut to 76 bp in-silico to mimic a 150-cycle sequencing kit. Results show that there are no issues identifying virus strains, and even shows a small boost in lower coverage (Table 5). High depth and coverage allow for accurate viral genome variant calling. Through this, any genomic changes that occurred could be identified, and consensus sequences could be generated based on these variations (Figure 6). Researchers can utilize these sequence data to figure out what changes were made and where the virus strain originated from.

Table 4. Mean depth and coverage of a clinical sample

Sequencing Cycle	Mapped Correctly	Sequencing Amount	Mean Depth	Coverage (1X)	Coverage (10X)	Coverage (100X)
Clinical 300 cycles	100.00%	710,316,080	4733.43	99.89%	99.86%	99.03%
Clinical 150 cycles	100.00%	357,510,080	3391.19	99.95%	99.87%	98.95%

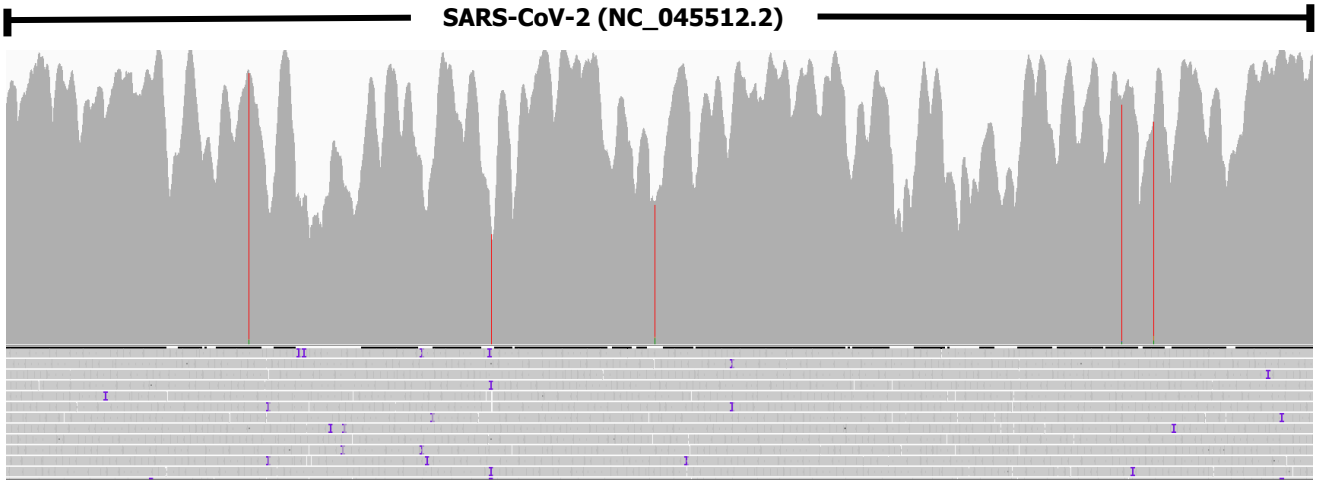

Figure 6. Variants found in clinical sample. Data was generated through the analysis of the identical sample above.

Summary

With the ever-growing threat of new global viral pandemics, the need for accurate identification and analysis of known and novel viruses have become more pressing than ever, both to prevent the occurrence of such outbreaks and expedite treatment of those affected by these diseases. Utilizing powerful NGS technology, the Celemics Comprehensive Respiratory Virus Panel (CRVP) aims to enable researchers in understanding and studying current and new viral strains that affect humans, and better equip them to research and create new treatments for patients. Through its validation through clinical samples and analytical validation of species differentiation, the CRVP Panel will act as a capable tool for deeper research into the sequencing and understanding of viral infections.

References

[1] Thorburn, F., Bennett, S., Modha, S., Murdoch, D., Gunson, R., & Murcia, P. R. (2015). The use of next generation sequencing in the diagnosis and typing of respiratory infections. *Journal of Clinical Virology, 69*, 96-100. doi:10.1016/j.jcv.2015.06.082