NGS Target Enrichment of Viral Pathogens using Twist Respiratory Virus Research Panel

INTRODUCTION

Accurately detecting and identifying viral pathogens is a critical global health concern exacerbated by the 2019 novel coronavirus (SARS-CoV-2) pandemic. A wide range of viral pathogens cause similar patient symptoms, making it difficult to identify the underlying infectious agent. Currently, RT-PCR assays are routinely used to detect viral pathogens. These common assays are rapid, but are often used to test for just one pathogen at a time. Multiplexed RT-PCR saves time, allowing for the simultaneous detection and identification of multiple viruses, but it suffers from other shortcomings, including lack of viral sequence information. Next generation sequencing (NGS) hybrid capture combines high-throughput capabilities with high sensitivity, making it possible to quickly identify specific, whole viral genomes from complex samples.

In the Twist SARS-CoV-2 Research Panel, we highlighted how NGS-based target enrichment was successfully applied for the detection and characterization of the SARS-CoV-2 viral genome. We showed that target enrichment resulted in a nearly million-fold enrichment of viral fraction post-capture for low viral titer samples and 99.9% genome coverage using only 25,000 reads for samples with a viral titer of 1 thousand copies. Moreover, we demonstrated the tolerance of the capture-based method to virus mutations and successfully identified various mutations in the SARS-CoV-2 virus. The ability to simultaneously detect and characterize viral genomes makes hybrid capture a powerful alternative to RT-PCR, that can monitor viral evolution and conduct population scale surveillance.

We've built upon Twist's expertise in utilizing target capture for infectious diseases by creating a panel that is capable of detecting multiple viral pathogens associated with common symptoms of respiratory illnesses. This panel, called the Twist Respiratory Virus Research Panel, is designed to enrich for 29 common respiratory viruses. To validate this panel, we synthesized 15 viral controls, both ssRNA and ssDNA, that were used in addition to the Twist SARS-CoV-2 Synthetic RNA control. These controls were spiked into human reference RNA to generate TruSeq-compatible DNA libraries for target enrichment. Here, we demonstrate the:

1. Capture of 16 different synthetic viral strains at 1,000,000 copies.
2. Capture of 3 viral strains across viral titers of 100–1,000,000 copies.
3. Simultaneous capture of two viral genomes at 10,000 copies each in simulations of co-infections.
4. Efficient multiplexed capture of diverse viral genomes at viral titers ranging from 100–1,000,000 copies.

RESULTS

The Twist Respiratory Virus Research Panel is targeted against the reference sequences for 29 common human respiratory viruses, including coronavirus (CoV), influenza virus, adenovirus, bocavirus (hBoV), enterovirus, metapneumovirus, parainfluenza (hPIV), human rhinovirus (HRV), measles morbillivirus (MeV), mumps virus (MuV), rubella virus, and respiratory syncytial virus (RSV) (Table 4). Additional probes were also designed to target diverse genomes representing major influenza A and B outbreaks since the year 2000, and to incorporate diversity from 77 additional rhinovirus strains.

To validate the performance of the Twist Respiratory Virus Research Panel, a variety of synthetic viruses were designed and synthesized as ssRNA or ssDNA, depending on their native genome structure. The synthetic viruses were sequence-validated, then diluted to a stock of ~1 million copies per ul for downstream experiments.

Each viral synthetic control was spiked into 50 ng of human reference RNA, which was then used to generate an Illumina TruSeq-compatible library (1,000,000 copies for each sample). The Twist Respiratory Virus Research Panel was then used to capture viral sequences, following the Twist Target Enrichment Protocol with a 16-hour hybridization time. In most cases, we found that over 70 percent of reads came from viral genomes in these libraries, representing at least a 2500-fold enrichment over the spiked-in content (Figure 1).

The one notable exception was for human bocavirus (hBoV). The human bocavirus genome is the smallest among those tested. Shorter templates will contribute less content to the library at equal titers. Indeed, while the total percent of viral reads is generally low for shorter templates, the fold-enrichment (which is normalized to template length) is generally higher. (Figure 1). Additionally, bocavirus has an ssDNA genome, while every other tested virus has an ssRNA genome. Since similar length ssRNA viruses show much higher enrichment, it's likely that lower yield for ssDNA during library preparation was what caused lower capture efficiency.

Despite differences in capture between bocavirus and the other ssRNA viruses, all synthetic controls were sequenced to high depth in great uniformity with Fold 80 Base Penalty in the range of 1.2 to 1.5. At 1 million sequenced reads, we find all templates to be covered at a median depth of 1500x, with at least 99% of bases covered to at least 30x depth (Figure 1), sufficient for confident variant calling and *de novo* assembly.

Next, we assayed the sensitivity of the Twist Respiratory Virus Research Panel by capturing three synthetic ssRNA viral genomes (H3N2, H1N1, and HRV) at various titer loads (100, 10,000 and 1,000,000 copies per library), using a 16-hour hybridization time. We found each virus to be enriched by at least 5000-fold at every tested titer, with greater than 20,000-fold enrichment for the viral template at low titers (Table 1). In summary, the Twist Respiratory Virus Research Panel can efficiently enrich for viral sequences present at titers spanning several orders of magnitude, demonstrating a limit of detection as low as 100 copies.

The Twist Respiratory Virus Research Panel is designed to target several viruses in a single capture, for the detection or characterization of several viruses in one reaction. Human respiratory pathogen co-infections often occur among patients suffering from respiratory distress and discomfort. Some common co-infections include respiratory syncytial virus (RSV) and coronaviruses (Martin et al 2012) or human bocavirus and parainfluenza viruses (Zhang et al 2012).

Co-infections were simulated by spiking in multiple synthetic viral controls into a single sample during library preparation. The following controls were spiked into 50ng human reference RNA at 10,000 copies of each virus per library:

- Human rhinovirus 89 (ssRNA) with Human bocavirus 1 (ssDNA)
- SARS-CoV-2 (ssRNA) with Human coronavirus 229E (ssRNA)
- SARS-CoV-2 (ssRNA) with Influenza H3N2 (ssRNA)
- SARS-CoV-2 (ssRNA) with Human rhinovirus 89 (ssRNA)

The libraries then underwent capture using the Twist Respiratory Virus Research Panel, using a 16-hour hybridization time. We assessed capture by measuring total viral content, fold-enrichment, and template coverage at different depths (Table 2). Both templates were completely covered at 1x in each co-infection experiment, with generally good uniformity across the template (Table 2 and Figure 2). This was true for viruses in the same family (such as the coronavirus 229E and SARS-CoV-2) as well as viruses in different families (such as human rhinovirus and human bocavirus). All the viruses in the experiment had >99% of the bases covered at 30x or above except human bocavirus. The relatively lower coverage at 30x and 100x depth for human bocavirus was likely due to lower library preparation efficiency from the ssDNA.

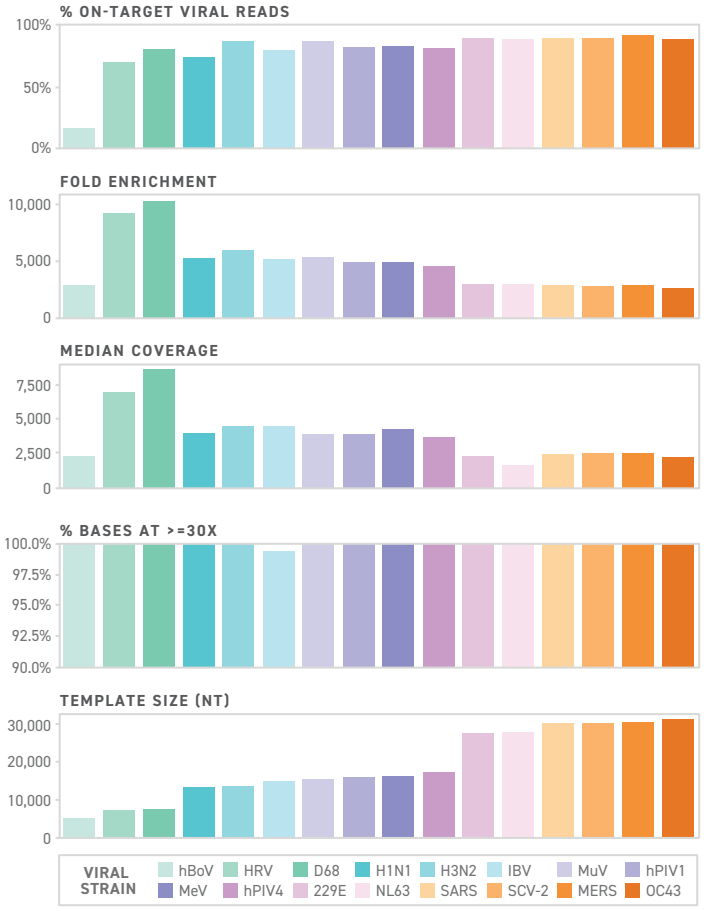

Figure 1: Detection of viral standards from different viral families. Each Twist synthetic viral control was spiked into 50 ng of human carrier RNA at 1,000,000 copies prior to cDNA synthesis. For each standard, we show percent of on-target viral reads, fold-enrichment above input, median coverage across the genome, the percent of bases with at least 30x coverage, and the template length. Viral genomes are ordered by template length (smallest to longest) among all genome segments. Full names for each virus abbreviation used are given in Table 3.

VIRAL STRAIN	VIRAL TITER (Copies)	READS EXPECTED WITHOUT ENRICHMENT (Out of 1M total)	NUMBER OF ON-TARGET VIRAL READS (Out of 1M total)	FOLD ENRICHMENT
H1N1	100	<1	402	28704
	10,000	1	53964	38531
	1,000,000	140	715054	5106
H3N2	100	<1	349	24025
	10,000	1	40081	27631
	1,000,000	145	818155	5640
HRV	100	<1	1453	190849
	10,000	1	26185	34393
	1,000,000	76	682086	8959

Table 1: Viral capture at different titer points using the Twist Respiratory Virus Research Panel. Synthetic viral standards for Influenza H1N1, Influenza H3N2, or HRV were spiked into a background of human RNA at 100, 10,000 and 1,000,000 copies. Expected and observed read counts from the viral template are shown, as well as fold-enrichment from capture. Numbers represent the mean of two replicates.The full names for each virus abbreviation are given in Table 3.

GROUP	VIRAL STRAIN	PERCENT ON-TARGET VIRAL READS	FOLD ENRICHMENT	1X COVERAGE	30X COVERAGE	100X COVERAGE
1	HRV	2.1%	27263	100.0%	99.4%	91.4%
	hBoV	0.2%	4218	100.0%	63.6%	0.0%
2	SCV-2	23.4%	73403	99.9%	99.9%	99.7%
	229E	6.8%	23380	99.9%	99.7%	90.9%
3	SCV-2	21.9%	68657	99.9%	99.9%	99.7%
	H3N2	3.3%	22678	100.0%	99.3%	86.6%
4	SCV-2	30.2%	94734	99.9%	99.9%	99.8%
	HRV	2.2%	29070	100.0%	99.8%	93.6%

Table 2: Detection of simulated co-infections. Two synthetic viral standards were spiked at 10,000 copies per virus into a background of 50ng of human RNA. Percent on-target reads, enrichment, and coverage at 1x, 30x and 100x depth are shown for both strains in each experiment. Numbers represent the mean of two replicates. Full names for each virus abbreviation are given in Table 3.

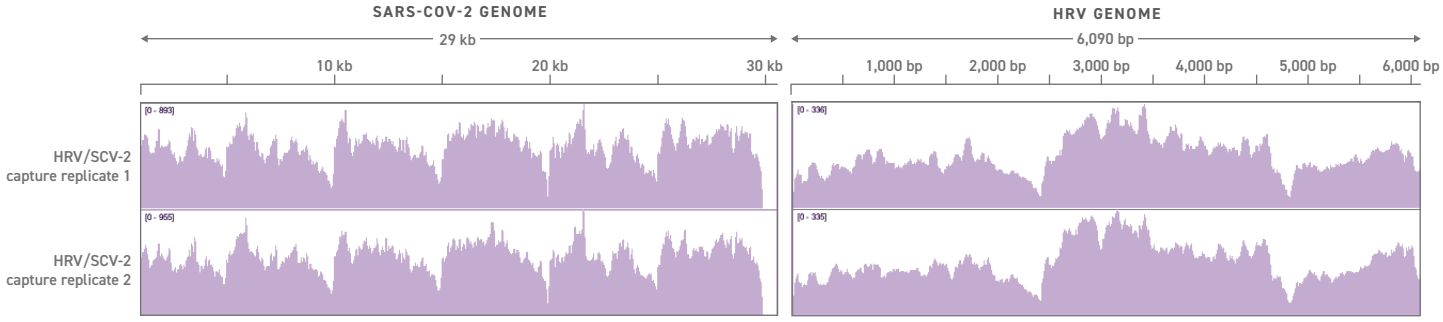

Figure 2: Genome browser views of read density from co-infection capture experiments using synthetic SARS-CoV-2 and human rhinovirus (HRV) standards. Full names for each virus abbreviation are given in Table 3.

Finally, we tested the Twist Respiratory Virus Research Panel in a multiplex capture system. This system allows investigators to decrease assay costs and time while maintaining equally efficient viral capture as compared to a single-plex capture. We demonstrate this by pooling 8 samples (with unique library indices) at various viral titers into a single hybridization tube, creating an 8-plex capture reaction.

In Figure 3, the first graph shows comparisons using a multiplex and single-plex hybridization reaction using 1,000,000 copies per library, while the second graph shows a similar comparison using 100, 10,000, and 1,000,000 copies per library. Each viral control was captured using the Twist Standard Target Enrichment workflow and the Twist Respiratory Virus Research Panel. This data demonstrates that 8-plex capture gives comparable efficient enrichment as a single plex capture when using the Twist Respiratory Virus Research Panel.

Due to the high capture efficiency of the Twist Target Enrichment Protocol and Twist Respiratory Virus Research Panel, extremely low levels of contamination between samples can sometimes be seen in the negative control data results. As a general note, to minimize risk of cross contamination between samples and/or controls, viral control and negative control libraries should be generated at different times, or in physically separated lab workspaces. If physical separation is not possible, cross contamination can be reduced by leaving an empty well between each sample during the library preparation process.

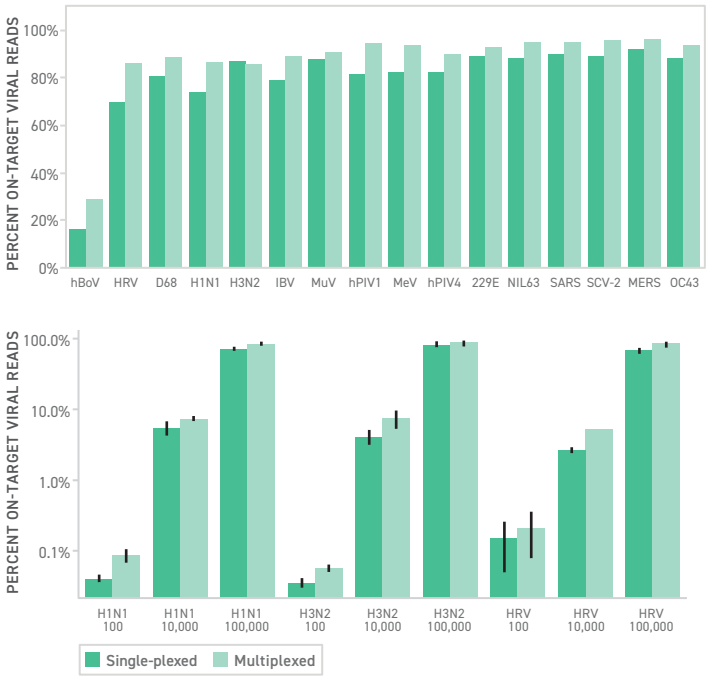

Figure 3. A comparison of the percent on-target viral reads captured using the Twist Synthetic Viral Controls at different viral titers, using a multiplex and single-plex hybridization reaction. Full names for each virus abbreviation are given in Table 3.

SUMMARY

Due to their common transmission patterns, respiratory pathogens are extremely contagious and outbreaks are difficult to control. The ability to monitor viral genome evolution and to characterize novel strains of these pathogens is crucial for containment and epidemiological studies that may inform public health decisions. The Twist Respiratory Virus Research Panel, along with the Twist Target Enrichment solution and Next Generation Sequencing, gives investigators the opportunity to collect meaningful data for multiple respiratory pathogens in a single assay. These workflows are compatible with a multiplexed system, decreasing cost and time without impacting the quality of sequencing results. The Twist Respiratory Virus Research Panel leverages a straightforward target enrichment workflow for detection and characterization of respiratory pathogens.

MATERIALS AND METHODS

Twist synthetic RNA and DNA viral controls were spiked into a background of 50 ng human reference RNA (Agilent) with viral copy numbers ranging from 100 to 1,000,000 copies per sample (Table 3). Co-infections were simulated by spiking multiple synthetic viral controls into one sample. A negative control consisting solely of human reference RNA was processed in parallel. The samples were then converted to single-stranded cDNA through random priming using NEB's Random Primer 6 (S1230S), ProtoScript II First Strand cDNA Synthesis Kit (E6560S). Single-stranded cDNA was then converted to dsDNA using the NEBNext Ultra II Non-Directional RNA Second Strand Synthesis kit (E6111S). The samples were converted to Illumina TruSeq-compatible libraries using Twist Library Preparation Kit using Enzymatic Fragmentation (PN 101059 and 100401) and Unique Dual Indices (UDI) (PN 101307).

Enrichment was performed using the Twist Respiratory Virus Research Panel (PN 103066, 103067 & 103068) and 500 ng of library in single-plex capture reactions using a 16-hour hybridization. For multiplexing experiments, 8 libraries were pooled (187.5 ng each) for a total of 1500 ng. Enriched libraries were sequenced with 2x75 bp paired-end reads on the Illumina NextSeq platform, using a NextSeq500/550 High Output kit. Alignment was performed with BWA against a custom genome index comprising the human genome (build hg38) concatenated with reference sequences for each virus in the panel. All data were downsampled to 1M mapped reads per sample, unless otherwise noted.

REFERENCES

Martin ET, Kuypers J, Wald A, and Englund JA. *Multiple versus single virus respiratory infections: viral load and clinical disease severity in hospitalized children*. Influenza Other Respir Viruses. 2012; 6(1): 71–77.

Zhang G, Hu Y, Wang H, Zhang L, Bao Y, and Zhou X. *High Incidence of Multiple Viral Infections Identified in Upper Respiratory Tract Infected Children under Three Years of Age in Shanghai, China*. PLoS One. 2012; 7(9): e44568.

CATALOG #	NAME	ABBREVIATED NAME	NUCLEIC ACID SPECIES	STORAGE
102023	Twist Synthetic SARS-CoV-2 RNA Control 2 (MN908947.3)	SCV-2	ssRNA	–90 to –70°C
103001	Twist Synthetic Influenza H1N1 (2009) RNA control	H1N1	ssRNA	–90 to –70°C
103002	Twist Synthetic Influenza H3N2 RNA control	H3N2	ssRNA	–90 to –70°C
103003	Twist Synthetic Influenza B RNA control	IBV	ssRNA	–90 to –70°C
103004	Twist Synthetic Human bocavirus 1 DNA control	hBoV	ssDNA	–90 to –70°C
103005	Twist Synthetic Human enterovirus 68 RNA control	D68	ssRNA	–90 to –70°C
103006	Twist Synthetic Human rhinovirus 89 RNA control	HRV	ssRNA	–90 to –70°C
103007	Twist Synthetic Mumps virus RNA control	MuV	ssRNA	–90 to –70°C
103008	Twist Synthetic Human parainfluenza virus 1 RNA control	hPIV1	ssRNA	–90 to –70°C
103009	Twist Synthetic Measles virus RNA control	MeV	ssRNA	–90 to –70°C
103010	Twist Synthetic Human parainfluenza virus 4 RNA control	hPIV4	ssRNA	–90 to –70°C
103011	Twist Synthetic Human coronavirus 229E RNA control	229E	ssRNA	–90 to –70°C
103012	Twist Synthetic Human coronavirus NL63 RNA control	NL63	ssRNA	–90 to –70°C
103013	Twist Synthetic Human coronavirus OC43 RNA control	OC43	ssRNA	–90 to –70°C

Table 3: Synthetic Controls used in validating the Twist Respiratory Virus Research Panel

VIRUS NAME	ACCESSION NUMBER
Human adenovirus 14	JN032132
Human adenovirus B1	NC_011203.1
Human adenovirus E	NC_003266
Human adenovirus type 7	AC_000018
Human bocavirus 1	MG953830.1
Human coronavirus 229E	NC_002645.1
Human coronavirus HKU1	NC_006577.2
Human coronavirus NL63	NC_005831.2
Human coronavirus OC43	NC_006213.1
Human enterovirus 68	NC_038308.1
Human metapneumovirus	NC_039199.1
Human parainfluenza virus 1	NC_003461.1
Human parainfluenza virus 3	NC_001796.2
Human parainfluenza virus 4	NC_021928.1
Human rhinovirus 3	NC_038312.1
Human rhinovirus 89	NC_001617.1
Human rhinovirus C	NC_009996.1
Human rubulavirus 2 (parainfluenzavirus 2)	NC_003443.1
Influenza B	NC_002211, NC_002204, NC_002210, NC_002209, NC_002208, NC_002207, NC_002206, NC_002205
Influenza H1N1 (2009)	NC_026432, NC_026431, NC_026434, NC_026436, NC_026433, NC_026437, NC_026435, NC_026438
Influenza H3N2	NC_007369, NC_007373, NC_007372, NC_007371, NC_007366, NC_007368, NC_007367, NC_007370
Measles	NC_001498.1
MERS	JX869059.2
Mumps	NC_002200.1
Respiratory syncytial virus (A)	NC_001803.1
Respiratory syncytial virus (B)	NC_001781
Rubella	NC_001545.2
SARS	NC_004718.3
SARS-CoV-2	NC_045512.2

Table 4: Viruses Targeted in Twist Respiratory Virus Research Panel

KEY COMPONENTS

CATALOG #	NAME	DESCRIPTION	STORAGE
101059: 16 rxn 101058: 96 rxn	Twist Library Preparation EF Kit Twist Library Preparation EF Kit 1 Twist Library Preparation Kit 2	Reagents for library construction 5x Fragmentation Enzyme 10x Fragmentation Buffer DNA Ligation Mix DNA Ligation Buffer Amplification Primers, ILMN (Tubes 100220, 100583 are not required when used with universal adapters) DNA Purification Beads	 –25 to –15°C 2 to 8°C
100401: 16 rxn 100573: 96 rxn	Twist Library Preparation Kit 2	DNA Purification Beads (as a Standalone Product, Bead Purification is also needed during cDNA Synthesis)	2 to 8°C
101307: 16 rxn 101308, 101309, 101310, 101311: 96 rxn	Twist Universal Adapter System - TruSeq Compatible	Twist Universal Adapters and Twist UDI Primers, provides unique dual-indexed combinations with 1 reaction per index pair	–25 to –15°C
103066: 2 rxn 103067: 12 rxn 103068: 96 rxn	Twist Respiratory Virus Research Panel & One Codex Software	Custom DNA Panel for Respiratory Viral Detection & One Codex software analysis credits	–25 to –15°C
100856: 2 rxn 100578: 12 rxn 100767: 96 rxn	Twist Universal Blockers	For the prevention of nonspecific capture: Universal Blockers Blocker Solution	–25 to –15°C
101262: 2 rxn 100983: 12 rxn 100984: 96 rxn	Twist Binding and Purification Beads	For target enrichment and purification: Streptavidin Binding Beads DNA Purification Beads	2 to 8°C
101279: 2 rxn 101025: 12 rxn 101026: 96 rxn	Twist Hybridization and Wash Kit (2 Boxes) Twist Hybridization Reagents (Box 1 of 2) Twist Wash Buffers (Box 2 of 2)	For target enrichment with standard hybridization: Hybridization Mix Hybridization Enhancer Amplification Primers Binding Buffer Wash Buffer 1 Wash Buffer 2	 –25 to –15°C 2 to 8°C