# NGS-Based Target Capture for SARS-CoV-2 Detection and Characterization

## INTRODUCTION

SARS-CoV-2, a single-stranded RNA virus, is the cause of the ongoing pandemic of Coronavirus Disease (COVID-19). Reverse Transcription PCR (RT-PCR) is a rapid method for detecting COVID-19 cases by using primers to amplify regions of interest from the SARS-CoV-2 virus. However, the need exists not only for rapid testing, but also highly sensitive tests capable of characterizing the entire viral genome, including any mutations, to monitor the viral spread and evolution over time.

Twist's nucleic acid hybridization capture-based assay has been used by researchers as a way of detecting and tracking emerging viral pathogens. Previously, Twist collaborated with the U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID) and Illumina to develop a large panel of capture probes called Pan Viral 1.0 for target enrichment for NGS of all human viruses. It includes >600,000 probes for target enrichment of >1,000 human viruses. The Pan Viral 1.0 NGS assay panel was used by USAMRIID to capture Ebola virus during the 2015 Western African outbreak, and by Institute Pasteur in Dakar, Senegal to characterize Human Monkeypox Virus during an outbreak in Nigeria in 2017.

During the ongoing COVID-19 pandemic, we used a similar but more specific approach for the detection and characterization of SARS-CoV-2. We designed a capture panel that includes probes to target the full SARS-CoV-2 (MN908947.3) reference genome. We also synthesized the full-length SARS-CoV-2 (MT007544.1) genome in six non-overlapping pieces and transcribed them into RNA. The SARS-CoV-2 (MT007544.1) genome contains 3 SNPS and one indel region compared to the panel genome. The synthetic viral ssRNA was spiked into human RNA reference as a positive control of the capture assay. We demonstrate here:

1. Detection of as low as single-digit copies of viral material with a fold enrichment of near a million;
2. Coverage of > 99.9% of the genome at 1X or greater after enrichment;
3. Identification of mutations in the SARS-CoV-2 control MN908947.3. .

## MATERIALS AND METHODS

To generate a synthetic RNA SARS-CoV-2 control, the reference genome of MT007544.1 (ncbi.nlm.nih.gov/nuccore/MT007544) and MN908947.3 (ncbi.nlm.nih.gov/nuccore/MN908947) was split into six non-overlapping fragments. Fragments were synthesized as double stranded DNA and then transcribed into RNA. The SARS-CoV-2 (MT007544.1) RNA control was subsequently used for capture.

The synthetic SARS-CoV-2 RNA was spiked into a background of 50 ng of human reference RNA (Agilent Technologies (740000-41)) with viral copy numbers ranging from single digit copies, 1, to 1,000,000 copies per sample (Table 1). A negative control consisting solely of human reference RNA was also processed in parallel. The RNA samples were then converted to cDNA through random priming using NEB's Random Primer 6 (S1230S), ProtoScript II First Strand cDNA Synthesis Kit (E6560S), and NEBNext Ultra II Non-Directional RNA Second Strand Synthesis kit (E6111S). The cDNA 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 with the Twist SARS-CoV-2 Research Panel (PN 102016, PN 102017, or PN 102018) using 500 ng of library in single-plex capture reactions using a 16-hr hybridization. Enriched libraries were sequenced with 2x75bp paired-end reads on the Illumina NextSeq platform using a NextSeq500/550 High Output kit. Alignment was performed with BWA to the SARS-CoV-2 genome sequence and the human reference genome (hg38). Aligned reads were downsampled to 1 million reads per sample, unless otherwise stated.

## RESULTS

The SARS-CoV-2 Research Panel was designed against the SARS-CoV-2 genome provided as MN908947.3 in RefSeq. As a positive control for capture, we generated synthetic RNA controls from two reference SARS-CoV-2 genomes. Here we captured MT007544.1, which contains 3 SNPs and 1 indel to the Twist SARS-CoV-2 Research Panel. These mutations were purposefully included in the synthetic control to evaluate target enrichment's tolerance to mutations.

Pre-capture viral fractions were calculated by dividing the mass of viral spike-in RNA by total input mass. Post-capture viral fractions were determined by downsampling to 1 million reads and dividing viral read counts by total reads without removing duplicates. Fold enrichment was then calculated by dividing post-capture viral fraction by pre-capture viral fraction (Table 1).

| VIRUS COPY NUMBER | VIRUS (PG) | VIRAL FRACTION PRE-CAPTURE | TOTAL READS | VIRAL READS (% UNIQUE) | VIRAL FRACTION POST-CAPTURE | FOLD ENRICHMENT |
|---|---|---|---|---|---|---|
| 1,000,000 | 20 | 0.04000000% | 1,000,000 | 977,796 (87%) | 97.8% | 2,444 |
| 1,000 | 0.02 | 0.00004000% | 1,000,000 | 241,173 (37%) | 24.1% | 602,933 |
| 10 | 0.0002 | 0.00000040% | 1,000,000 | 3,506 (18%) | 0.351% | 876,500 |
| 1 | 0.00002 | 0.00000004% | 1,000,000 | 394 (33%) | 0.039% | 985,000 |
| Negative control | 0 | 0.00000000% | 1,000,000 | 26 (24%) | 0.003% | N/A |

**Table 1:** Detection of viral reads at different spike-in copy numbers into human RNA.

Viral reads were detected in all positive samples. Enrichment ranged from 2,444X to 985,000X and was inversely correlated to initial spike-in concentration (Table 1). The high-duplication rate, seen by the low number of unique reads at low viral titers, was not unexpected given the extremely limited amount of starting material. These results indicate that viral RNA can be detected with as few as a single-digit number of copies using as low as 25,000 reads (Table 2). We observed a small amount of viral material in the negative control, highlighting the importance of physically isolating samples when working with extremely sensitive assays. As the level of contamination was more than 10-fold lower than what was observed with a single copy, we are confident that the positive result for a single copy is not simply due to background.

| | PERCENT OF GENOME COVERED AT 1X | | | | | |
|---|---|---|---|---|---|---|
| VIRUS COPY NUMBER | 25K READS | 100K READS | 200K READS | 500K READS | 1M READS | 8M READS |
| 1,000,000 | 99.9% | 99.9% | 99.9% | 99.9% | 99.9% | 99.9% |
| 1,000 | 99.9% | 99.9% | 99.9% | 99.9% | 99.9% | 99.9% |
| 10 | 17.4% | 51.2% | 69.5% | 83.2% | 86.5% | 91.5% |
| 1 | 1.4% | 7.8% | 16.9% | 21.7% | 25.3% | 27.2% |
| Negative control | 0.0% | 0.0% | 0.7% | 1.3% | 1.3% | 1.4% |

**Table 2:** Coverage of SARS-CoV-2 genome at different viral copy numbers with varying total reads.

At moderate viral titers (1,000 copies), we were able to recover 1x coverage across the entire genome (with the exception of the endogenous poly-A tail) with only 25,000 sequenced reads per sample. Even at low titers (10 copies), 83% of the genome was covered at least 1x with 500k reads and over 90% of the genome was covered at least 1x depth with 8M reads per sample (Table 2). While a small portion (up to 1.4% with 8M reads) of the genome was covered in the negative control, our capture with a single copy of the genome was substantially higher (up to 27%), indicating that capture at this titer does not reflect background contamination between samples.

Variants were captured with high efficiency. MT007544.1 contains three single nucleotide substitutions compared to the panel genome, all of which were detected by over 99% of mapping reads. For the 10 nucleotide deletion in MT007544.1, we found that 99% of reads spanning the deletion site correctly detected the deletion (Figure 1).

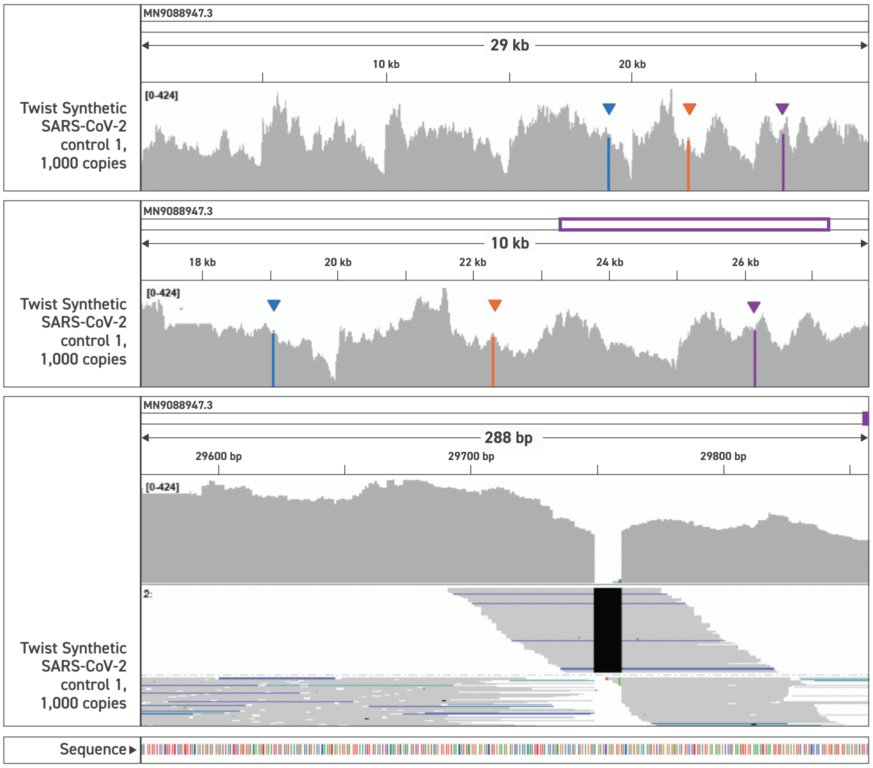

**Figure 1:** Detection of variants present on the synthetic control based on MT007544.1 reference. Single nucleotide substitutions in the top and center panels are indicated with triangles. The bottom panel highlights a known deletion between the panel genome and MT007544.1 synthetic control.

## DISCUSSION

We show here target capture is highly sensitive, being able to detect even single-digit copies of the virus. In addition, we are able to recover the full virus genome sequences, facilitating phylogenetic analysis and enabling studies into the lineage and evolution of the SARS-CoV-2 virus. Moreover, we demonstrated the tolerance of the capture-based method to virus mutations and successfully identified various mutations. The ability to simultaneously detect and characterize the SARS-CoV-2 virus makes the capture-based method not only a powerful alternative to the RT-PCR based method, but also an invaluable tool for monitoring viral evolution and development and for population-scale surveillance.

SARS-CoV-2 is of particular interest, but the principles of panel design and capture can be applied to broader targets and applications. For example, we developed a respiratory disease panel (PN 103066, 103067 and 103068), and the Comprehensive Viral Research Panel. The respiratory-disease panel covers a number of respiratory pathogens such as influenza, coronavirus, rhinovirus, adenovirus and so on, and could help identify and differentiate respiratory pathogens with similar clinical symptoms. The Comprehensive Viral Research Panel with updated content and workflow to target all human viral pathogens can enable the infectious disease community to further detect and quickly characterize emerging threats.

## KEY COMPONENTS

| PART NUMBER | NAME | STORAGE |
|---|---|---|
| 102019 | Twist Synthetic SARS-CoV-2 RNA Control 1 (MT007544.1) | −90 to −70°C |
| 102024 | Twist Synthetic SARS-CoV-2 RNA Control 2 (MN908947.3) | −90 to −70°C |
| 102016: 2 rxn<br>102017: 16 rxn<br>102018: 96 rxn | Twist SARS-CoV-2 Research Panel | −25 to −15°C |

*The following trademark representations are believed to be accurate, but are not guaranteed to be so: Agilent Technologies®, Illumina®, NEB®, NEBNext®, NextSeq®, ProtoScript®, and TruSeq®. Registered or registration-pending names, trademarks, etc used herein, even when not specifically marked as such, are not to be considered unprotected by applicable law.*

*These products are for research use only, and subject to additional use restrictions as set forth in Twist's Supply Terms and Conditions: www.twistbioscience.com/supply-terms-and-conditions*

DOC-001158 REV 3.0 | LAST REVISED: February 22, 2021

| REVISION | DATE | DESCRIPTION |
|---|---|---|
| 3.0 | Feb 22, 2021 | Page 3: Replaced Pan Viral 2.0 with Comprehensive Research Panel. |