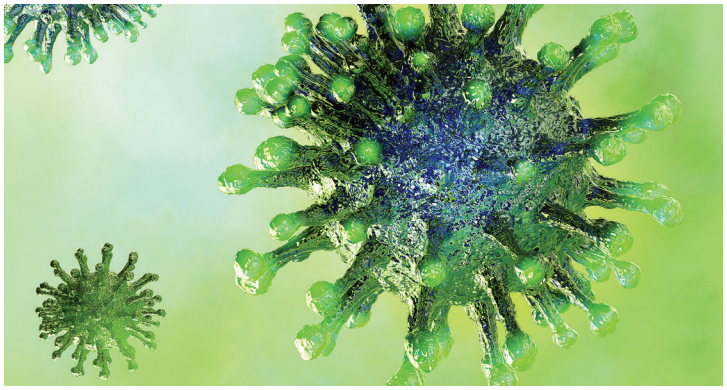
GenapSys™ Sequencing Platform:

SARS-CoV-2 Genome Sequencing using an Amplicon Panel

- **Decentralized and low cost NGS solution for surveillance and testing for the Covid-19 pandemic**
- **Accurate detection of SARS-CoV-2 variants, with uniform genome coverage**
- **Rapid library prep assay that can be completed in less than 6 hours**
- **Detection of a wide range of viral titers (5-orders of magnitude) and good performance at low read coverages**
- **Multiplex samples to enable higher throughput and a lower cost per sample**



Introduction

SARS-CoV-2, is a single-stranded RNA virus. Detection of the virus in a sample is typically done with Reverse Transcription PCR (RT-PCR) which utilizes primers to amplify a specific region of interest from the SARS-CoV-2 genome and provides a rapid response. Equally important to performing rapid testing is the need to track the spread and mutational evolution of the viral genome in an accurate and distributable manner.

Next-Generation Sequencing (NGS) provides an effective and comprehensive method to interrogate the full genome of the virus and its mutations vs. the limited focus of current testing methodologies. Epidemiological studies require more information than just the presence or absence of the virus, and a complete view of the genome across a large number of patient samples can provide the ability to monitor transmission and viral evolution. NGS is a powerful tool that provides a comprehensive view of the viral genome and allows the ability to correlate the host genotype and immune response to the virus. Thus, it enables essential insights into viral function and evolution and can help better direct therapeutic research and vaccine development efforts.

Technology

The GenapSys™ Sequencer employs a novel electrical detection method that is capable of generating highly accurate DNA sequence information. With a CMOS based detector, simple fluidics, and low computational requirements, the GenapSys instrument is small, affordable, and accessible even to novice genomic scientists. Inside the sequencing chip are millions of individual sensors, each loaded with a single bead coated in thousands of clonal copies of an individual DNA template. Individual nucleotides are flowed across the chip in succession and successful incorporation is detected by changes in electrical impedance as the complementary DNA strand grows. These impedance changes are converted into high accuracy base calls and the results are output as a FASTQ file.

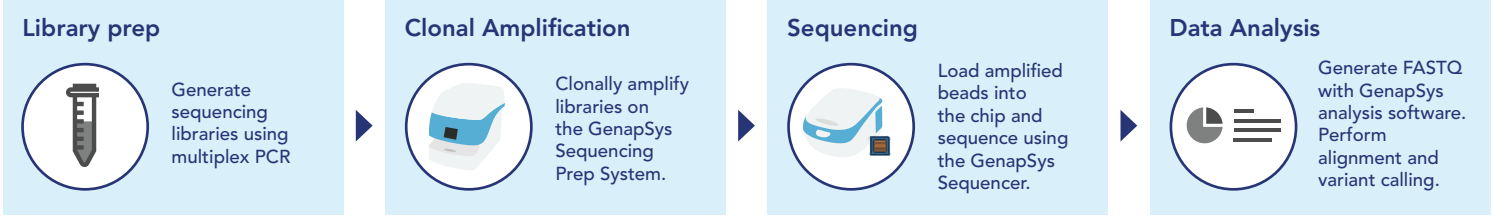
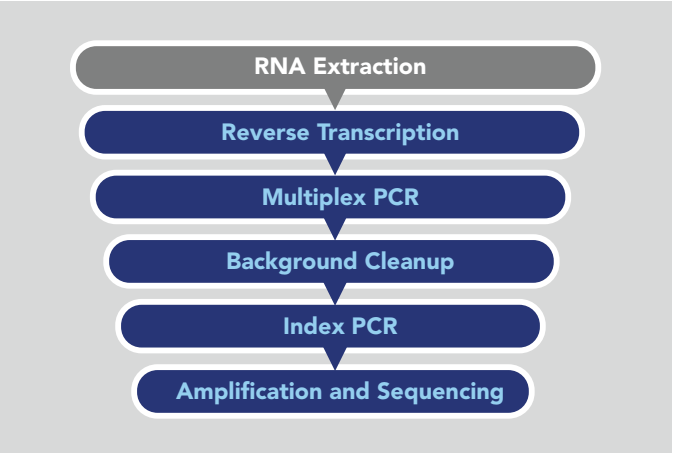


Fig. 1 The GenapSys workflow for the SARS-CoV-2 assay starts with RNA extracted from patient samples and then cDNA synthesis followed by library preparation using a Multiplex PCR panel. The libraries are clonally amplified on the GenapSys Sequencing Prep System and are sequenced using the GenapSys Sequencer. The results consist of FASTQ files for each sample and variant calls using the GenapSys data analysis pipeline.

Methods

GenapSys has demonstrated high accuracy sequencing of the SARS-CoV-2 genome using a multiplex amplicon PCR panel. The multiplex PCR panel consists of two pools of primer sets that result in a total of 343 tiled overlapping amplicons, that cover the SARS-CoV-2 genome. Here, we highlight the performance of the multiplex amplicon PCR panel for SARS-CoV-2 on the GenapSys Sequencing platform (Table 1), over a wide range of viral loads (5 orders of magnitude) and reads per sample (3 orders of magnitude). Briefly, two synthetic SARS-CoV-2 RNA controls that differ by 4 variants were spiked into Universal Human Reference RNA (UHRR) at different viral copy numbers. The RNA samples were converted to cDNA through a reverse transcription step, followed by multiplex PCR amplicon library generation from the viral genome. The libraries were clonally amplified on the GenapSys Sequencing Prep instrument and sequenced using the GenapSys sequencer, as shown in Fig. 1. The FASTQ files generated were analyzed for genomic coverage performance and detection of variants at different reads per sample.



Assay performance was demonstrated using two synthetic RNA SARS-CoV-2 control standards generated from Twist Bioscience, which were based on the reference genome of MN908947.3 (ncbi.nlm.nih.gov/nuccore/MN908947) and MT007544.1 (ncbi.nlm.nih.gov/nuccore/MT007544) (Catalog #102024 and 102019 respectively). MT007544.1 differs from MN908947.3 (Refseq sequence) by 4 variants (3 SNVs and a 10 bp deletion). Each standard is split into six non-overlapping fragments, and quantitated to 1 million viral genome copies/µL. Assay performance across a wide range of viral titers was tested by spiking in as many as 1 million copies and as low as 10 copies of each standard into 50 ng of UHRR (Agilent Part #740000). For real samples, we recommend that customers evaluate commercially available kits capable of extracting viral RNA. Additional information on compatible kits for viral RNA extraction can be found here: https://www.fda.gov/media/134922/download.

The RNA mix was converted to cDNA, using a reverse transcription step involving random primers using reagents provided in the Genapsys SARS-CoV-2 Panel Kit (see ordering information on page 6). The cDNA was split into two parallel PCR reactions performed with the two pools of primers. The PCR reactions were pooled together, cleaned up and a background cleanup digestion reaction was performed to remove non-specific PCR products. A second round of PCR was performed to attach indexing oligos followed by a final cleanup. Library quantification is performed prior to mixing different samples. Single sample or multiplexed

Genapsys SARS-CoV-2 Panel Specifications	
Parameter	**Specification**
Enrichment Method	Multiplex PCR
Strain Compatibility	Complete coverage of major strains: MN908947 and MT007544
Cumulative Target Size	29,903 bp
Number of Amplicons	343
Amplicon Size	116 - 196 bp, median 149 bp
Number of Primer Pools	2
Sample Input Requirement	5 - 11 uL of extracted total RNA; compatible with range of RNA inputs, nominally 50 ng of purified total RNA
Sample Types	Sputum, nasopharyngeal and oropharyngeal swabs or aspirates, tissue, and other sources of viral RNA
Total Assay Time	< 6 hours
Hands on Time	< 1 hour
Number of Samples per Run (≥ 50K reads / sample)	192
Number of Samples per Run (≥ 100K reads / sample)	96
Number of Samples per Run (≥ 330K reads / sample)	32

To learn more about the GenapSys Sequencing Platform, visit GenapSys.com

sample or multiplexed libraries were clonally amplified using the GenapSys Sequencing Prep instrument, and were then sequenced on the GenapSys Sequencer, using the protocol in the GenapSys Sequencer User Guide.

BWA MEM was used to align reads to the SARS-CoV-2 genome sequence and the human reference genome (hg38). Sequencing runs with multiple samples are demultiplexed, and the FASTQ sequence and a SARS-CoV-2 whole genome sequencing report are generated per sample. The report specifies the number of viral reads, depth of coverage and variant calls with reference to the Genbank (MN908947)/ RefSeq (NC_045512.2) sequence. Variant calling was performed using an internally developed analysis pipeline, based on Vardict.

Results

We first demonstrated the enrichment capability of the GenapSys SARS-CoV-2 Panel. We generated libraries and sequenced samples with 1M, 1000, 100 or 10 viral copies of the SARS-CoV2-MT007544.1 control, or with 1000 copies of the SARS-CoV2-MN908947.3 control. Sequencing reads were mapped to the Human hg38 genome and the SARS-CoV-2 genome, and the enrichment was calculated as the ratio of the reads aligned to the SARS-CoV-2 genome divided by the total aligned reads. As shown in Fig. 2, this panel shows very high enrichment of the viral reads using this assay, with >99.6% of all reads mapping to the SARS-CoV-2 genome for samples ranging from as low as 10 viral copies to as high as 1M viral copies (5 orders of magnitude).

Next, we demonstrated the genomic coverage of the SARS-CoV-2 viral genome across 5 orders of viral load and across different read depths using this workflow. We observed that the genome coverage at 1x was ≥98% for samples with ≥1000 copies for two different read depths: 100K reads/ sample and 400K reads/sample (Fig. 3). The genome coverage at 1x was ≥96% for samples with as low as 10 viral genome copies, for both 100K and 400K reads per sample. For variant calling applications that require higher coverage, we demonstrated that the coverage at 30X was greater than ~94% for all samples with ≥100 copies for both read depths of 100K and 400K reads/ sample. This demonstrates the ability of this workflow to generate high coverage of the SARS-CoV-2 genome at low reads per sample, thus enabling effective multiplexing of large numbers of samples (see Specifications table). The theoretical sample multiplexing calculations are based on the genome coverage (at 1x) and assuming 50K, 100K or 330K reads/ sample, given 12M reads per sequencing run. In particular, ≥ 330K reads/ sample corresponds to the ability to run 32 samples per G3 chip and ≥ 100K reads/ sample corresponds to effective multiplexing of 96 samples per G3 chip. The low cost per sequencing run of the GenapSys platform and the high sample multiplexing results in low cost per sample.

The GenapSys SARS-CoV-2 Panel workflow has important applications in research, surveillance and detection of the SARS-CoV-2 virus. For these applications, it is critical to be able to identify variants in the viral genome sequence with

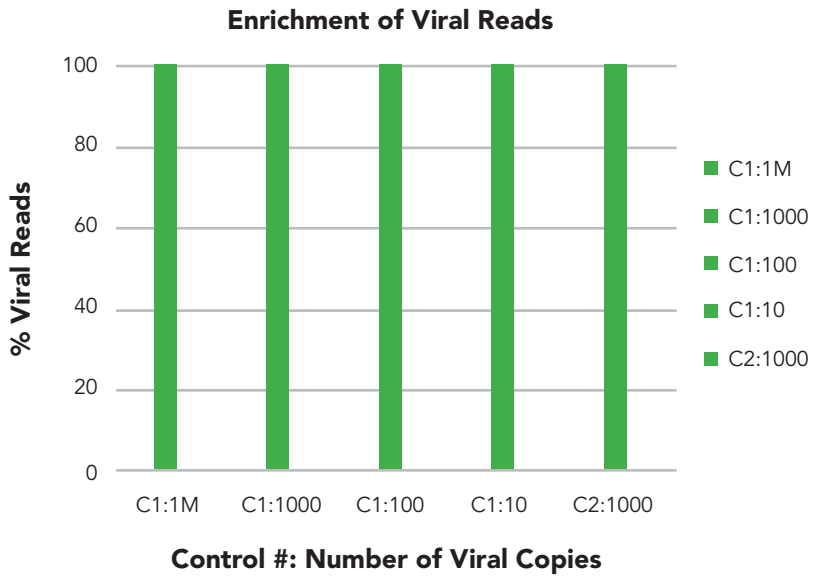


Fig. 2 High enrichment efficiency of the GenapSysSARS-CoV-2 Amplicon Panel workflow. Samples with viral copies ranging from 10 copies to 1M copies in 50 ng of UHRR, were enriched using the GenapSys multiplex PCR amplicon panel and workflow described in the Methods section. Greater than 99% of the reads map to the viral genome, indicating high enrichment.

To learn more about the GenapSys Sequencing Platform, visit GenapSys.com

high confidence and across different viral titers and reads per sample. We demonstrated the variant calling and genome coverage performance of the workflow for different read depths, using the sample for the sample containing 1000 copies of the MT007544.1 control. The MT007544.1 control has 4 variants: 3 SNVs and a 10 bp deletion as compared to the RefSeq sequence. As shown in Table 1, we were able to detect all 4 variants across a wide range of reads depths, from ~7.5M reads/ sample to as low as 50K reads/ sample. Additionally, the genome coverage at 1x and 30x of the sample is high, even at low read depths such as 50K - 100K reads/ sample. Finally, as shown in Table 2, the GenapSys SARS-CoV-2 Amplicon Panel workflow enables high accuracy variant calling across a wide range of viral inputs. The ability to accurately call variants across several orders of magnitude of viral input and read depth, highlights the high accuracy of the GenapSys platform, the coverage of the multiplex PCR amplicon pools, and the ability to multiplex large numbers of samples.

The GenapSys workflow for SARS-CoV-2 whole genome sequencing produces a FASTQ file as well as a report for each sample (see Fig. 4). The report references the number of reads mapped to the SARS-CoV-2 genome and the Human reference genome (hg 38), the mean read depth, the genome coverage at ≥1X, and the variants detected. The report lists the variants with reference to the RefSeq (NC_045512.2) sequence, and the depth and frequency of the variant calls.

Conclusion

The Genapsys SARS-CoV-2 Amplicon Panel workflow solution provides a decentralized, low-cost and accurate NGS solution for whole genome sequencing of the virus and detection of variants along with high sample multiplexing. The ability to have ready access to such genome-wide data across different geographies can provide public health experts with the ability to track the viral evolution, perform surveillance, and develop solutions to manage the current SARS-CoV-2 pandemic and monitor future viral outbreaks. Additionally, the GenapSys SARS-CoV-2 workflow can be combined with sequencing of patient gene variants and immune response genes, to highlight the underlying causes behind the wide variation in severity of Covid-19 response in patients. This capability will enable a broad range of academic and biopharma labs who are performing vaccine research and developing therapeutic solutions.

Fig. 3 Detection of SARS-CoV-2 virus with high genome coverage across 5 orders of viral loads. Samples were generated with 1M, 1000, 100 and 10 viral genome copies of the MT007544.1 control sample, and 1000 viral genome copies of the MN908947.3 control sample, spiked into 50 ng of UHRR respectively. Genome coverage ≥1X and ≥30X were high at both 100K reads and 400K reads/ sample. Viral copies as low as 10 copies were detected by the assay.

Genome Coverage and Variant Calling : C1- MT007544.1 control: 1000 viral copies					
Number of Reads	>7M	1M	250K	100K	50K
% Fraction of Viral Reads in Aligned Reads	99.99	99.99	99.99	99.99	99.99
% of genome represented at 1X	98.5	98.3	98.05	98.02	97.63
% of genome represented at 30X	98.2	98.2	95.94	94.78	92.84
Number of Variants called	4/4	4/4	4/4	4/4	4/4

Table 1: High genome coverage and accurate variant calling at low read depth using the GenapSys SARS-CoV-2 Amplicon PCR panel. The table shows sequencing metrics of the MT007544.1 control sample, with 1000 viral genome copies spiked into 50 ng of UHRR. The original sample sequencing data with >7M reads is downsampled to 1M, 250K, 100K and 50K sensors, for further analysis. The fraction of viral reads out of the aligned reads was > 99.9% for all read depths. Genome coverage at ≥1X was >97.5% at all read depths, and genome coverage at ≥30X was ≥94.7% at >100K reads. The genome coverage is not 100% in this sample, since the MT007544.1 synthetic control is present as 6 RNA fragments, and amplicons at the edges of the fragments may not be generated during the workflow. Importantly, all 4 variants were detected for the sample, even at low numbers of reads.

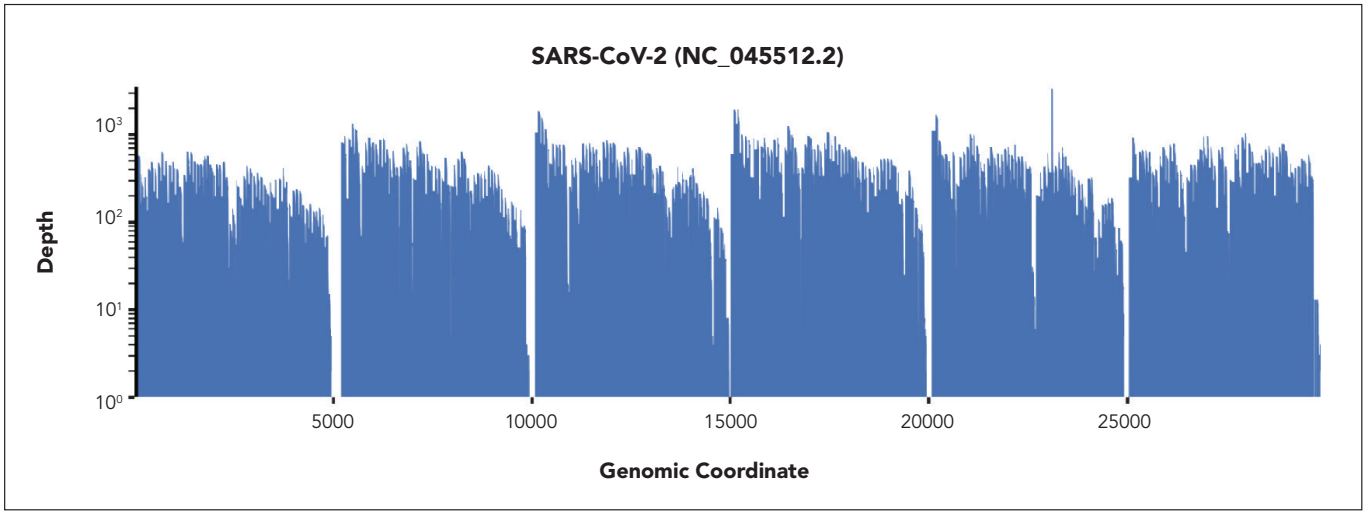
To learn more about the GenapSys Sequencing Platform, visit GenapSys.com

Frequency of Variant for different Viral Copy Inputs: C1 - MT007544.1 control				
Position	Variant	1M	1000	100
19065	T -> C	98	98	98
22303	T -> G	99	99	99
26144	G -> T	99	99	99
29749	ACGATCGAGTG -> A	100	100	100

Table 2: Accurate variant calling for different viral copy samples. Samples with 1M, 1000, 100 copies of the MT007544.1 control were processed with the GenapSys SARS-CoV-2 Amplicon PCR workflow. Each sample had ~100K reads or greater, and all 4 variants (3 SNVs and a 10bp deletion) were detected with high confidence, as demonstrated by the frequency of the variant in the reads (>97%).

SARS-CoV-2 (COVID-19) Sequencing Report (20200520.172734)

The sample contained 10 reads that mapped to the human genome (HG38) and 94,479 reads that mapped to the SARS-CoV-2 genome. With a mean depth of 403x, reads cover 98% of the SARS-CoV-2 genome. A total of 4 variants were detected.



Position	Variant	Depth	Frequency
19065	T -> C	505	97%
22303	T -> G	416	98%
26144	G -> T	489	99%
29749	ACGATCGAGTG -> A	13	100%

Fig. 4 Overview of the whole genome sequencing sample report generated by the GenapSys SARS-CoV-2 Multiplex PCR Amplicon panel workflow.

Ordering Information

The GenapSys SARS-CoV-2 Panels contain GenapSys Multiplex PCR Primers and GenapSys Targeted Library Kit with RT reagents. GenapSys Unique Dual-Indexed PCR Primers can be ordered separately to complete the workflow from input RNA to sequencing ready NGS libraries. For detailed product configurations, please refer to the GenapSys SARS-CoV-2 Panel User guide.

Product	Part Number
GenapSys™ 8-plex SARS-CoV-2 Kit (8 reactions)	1002681
GenapSys™ SARS-CoV-2 Panel (96 reactions)	1002633
GenapSys™ 32-plex Unique Dual-Indexed PCR Primers Set A and Set B (64 reactions)	1002685
GenapSys™ 32-plex Unique Dual-Indexed PCR Primers Set A and Set B (192 reactions)	1002686
GenapSys™ 96-plex Unique Dual-Indexed PCR Primers (96 reactions)	1002684

GENAPSYS, INC.
200 Cardinal Way, 3rd Floor, Redwood City, CA 94063 • (800) 796-7710 • GenapSys.com

GenapSys
Sequencing without compromise