frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 24 February 2023
DOI 10.3389/fgene.2023.1086865

OPEN ACCESS

EDITED BY
Gary S. Stein,
University of Vermont, United States

REVIEWED BY
Haridha Shivram,
Genentech Inc., United States
Yang Pan,
Beijing Center for Disease Prevention and Control (Beijing CDC), China

*CORRESPONDENCE
Lian Chye Winston Koh,
Winston_Koh@ibb.a-star.edu.sg
Shawn Hoon,
shawn@talolabs.com

SPECIALTY SECTION
This article was submitted to RNA, a section of the journal Frontiers in Genetics

RECEIVED 01 November 2022
ACCEPTED 25 January 2023
PUBLISHED 24 February 2023

CITATION
Koh LCW, Seow Y, Kong KW, Lau MLL, Kumar SK, Yan G, Lee CK, Yan B, Tambyah PA and Hoon S (2023), Sub genomic analysis of SARS-CoV-2 using short read amplicon-based sequencing. *Front. Genet.* 14:1086865. doi: 10.3389/fgene.2023.1086865

Sub genomic analysis of SARS-CoV-2 using short read amplicon-based sequencing

Lian Chye Winston Koh[1,2]*, Yiqi Seow[2,3], Kiat Whye Kong[3], Ming Li Lalita Lau[3], Shoban Krishna Kumar[4], Gabriel Yan[5], Chun Kiat Lee[6], Benedict Yan[6], Paul Anantharajah Tambyah[5] and Shawn Hoon[3]*

[1]Bioinformatics Institute, Agency for Science Technology and Research, Singapore, [2]Institute for Bioengineering and Bioimaging, Agency for Science Technology and Research, Singapore, [3]Institute for Molecular and Cell Biology, Agency for Science Technology and Research, Singapore, [4]Division of Advanced Internal Medicine, Department of Medicine, National University Health System, Singapore, [5]Division of Infectious Diseases, Department of Medicine, National University Health System, Singapore, [6]Molecular Diagnosis Centre, Department of Laboratory Medicine, National University Health System, Singapore

The novel coronavirus disease 2019 (COVID-19) pandemic poses a serious public health risk. In this report, we present a modified sequencing workflow using short tiling (280bp) amplicons library preparation method paired with Illumina's iSeq100 desktop sequencer. We demonstrated the utility of our workflow in identifying gapped reads that capture characteristics of subgenomic RNA junctions within our patient cohort. These analytical and library preparation approaches allow a versatile, small footprint and decentralized deployment that can facilitate comprehensive genetics characterizations during outbreaks. Based on the sequencing data, Taqman assays were designed to accurately capture the quantity of subgenomic ORF5 and ORF7a RNA from patient samples and demonstrated utility in tracking subgenomic titres in patient samples when combined with a standard COVID-19 qRT-PCR assay.

KEYWORDS

genomics, SARS-CoV-2, subgenomic RNA, transcriptome, temporal

1 Introduction

SARS-CoV-2 patients present with a range of severity and it is unclear how infectious asymptomatic cases are although they are believed to contribute to the spread of the virus. The proportion of asymptomatic patients have been estimated to range from 40%–45% and tend to be younger and do not have preexisting conditions (Oran and Topol, 2020). Conversely, there are many patients who remain PCR positive for weeks despite clinically recovering from the infection. This has posed an important question from a public health management perspective: When can we discharge or remove from isolation patients who are clinically well but remain PCR positive? Increasingly, the consensus is that PCR tests are picking up viral debris shed after an infection and that a time-based discharge criterion may be a more appropriate measure. WHO currently recommends that patients who are clinically well can be discharged 10 days after their symptoms began as long as they have been symptom-free for three consecutive days. Most countries now release patients from isolation between 7 and 14 days from onset of symptoms (He et al., 2020; MIN COW, 2020). In Singapore, COVID-19 patients who are clinically well are discharged at day 21 of their illness without the need for repeat PCR testing, though immunocompromised patients still require two negative PCR tests before discharge (Ministry of Health, 2022).

Coronaviruses express 3' proximal ORFs through subgenomic RNAs (sgRNAs) that are generated by a discontinuous transcription mechanism that is unique to RNA viruses (Sola et al., 2015). A study from Germany showed that infectious virus was isolated from throat or lung but not from stool samples and this correlated with evidence of active replication in throat samples by measuring the levels of transcribed subgenomic mRNA which are only produced by infected cells (Wölfel et al., 2020). The present study examines the expression of subgenomic RNAs in 53 patients. In addition, we demonstrated the ability of our workflow to capture temporal dynamics of amplicon coverage and subgenomic RNA junctional reads during an infection time course over 15 days.

2 Materials and methods

2.1 Patient samples

53 patients admitted to the National University Health System, from Feb to May 2020, with SARS-CoV-2 confirmed by RT-PCR (MiRXES, Singapore [Feb-Mar]; Cobas SARS-CoV-2 Roche, Switzerland [from Apr onwards]) from nasopharyngeal swabs were included in this study. Ethics approval was obtained from the National Healthcare Group Institutional Review Board.

2.2 Reverse transcription and quantitative PCR

400 µl of nasopharyngeal swab solubilized in UTM was extracted using EZ1 virus mini kit (Qiagen) into 90 µl elution buffer. An aliquot was heat-inactivated at 60°C for 30 min as required by Ministry of Health (Singapore) to prevent SARS-COV-2 infection and 6.5 µl of the heat-inactivated nucleic acid was reverse transcribed using SuperScript III (Life Technologies) (2 µl 5x FS buffer, 0.5 µl dNTPs, 0.5 µl 0.1M DTT, 0.25 µl SuperScript III) using 0.25 µl of Random Primer mix (NEB) using the following protocol: 65°C 20s, 4°C 60s, 55°C 50 min, 95°C 5 min.

2.3 Pooled PCR and library preparation

1 µl of the cDNA was then used for each pooled PCR in the following 10 µl Q5 polymerase reaction mix (NEB) (1 µl 10x Q5 buffer, 0.8 µl 10 mM dNTPs, 0.5 µl 10 µM pooled primers (Pool 1 or 2), 0.1 µl Q5 polymerase) using cycling protocols of 98°C 15s, 63°C 3 min s, 38 cycles). The two pools were combined after PCR and purified using 18 µl AmpureXP beads (Beckman-Coulter) as per manufacturer's protocol and eluted in 20 µl water. 10ng of the amplified product was prepared for library with ligation and barcoded using the NEXTFLEX Rapid DNA Seq 2.0 kit (Perkins-Elmer) as per manufacturer's instructions and eluted in 25 µl of water. Samples were pooled at 4-6 samples per run and diluted to 50pM before loading into the iSeq100 cartridge (150 nt read, paired end; Illumina) as per manufacturer's instructions.

2.4 Read processing, alignment and quality control

The critical step in the analysis is the alignment of partial transcript reads spanning sub genomic junctions to the reference genome. To achieve this, we developed a gapped alignment workflow that combines GSNAP (Wu and Watanabe, 2005) and V-ASAP (Maurier et al., 2019) to identify and quantify reads spanning these sub-genomic junctions. GSNAP was deployed over other available tools such as TOPHAT because it allows for an unbiased approach by not requiring an existing annotation of the subgenomic transcripts. Following alignment with GSNAP, gapped reads were filtered from the resulting SAM files based on CIGAR (Compact Idiosyncratic Gapped Alignment Report) strings using an AWK script. The remaining non-gapped reads are processed using V-ASAP, which merges the overlapping paired end reads and quality controls the merged amplicon reads to specifically those containing our intended forward and reverse primer sequences. Merged amplicon reads that passed quality control is realigned to the genome using BWA (Li and Durbin, 2009), this is followed by variant calling using SAMTOOLS and BCFTOOLS (Li et al., 2009).

The filtered gapped read SAM entries from GSNAP alignment, were further narrowed down to those spanning the known canonical subgenomic regions using RSAMTOOLS in R (Bioconductor, 2022). The quantity of these reads is then normalized by sequencing depth for comparison across the sample cohort and time points. Visualization of the downstream characterization are plotted using GGPLOT2 (Wickham, 2009). Reads Counts for subgenomic reads are represented in terms of reads per million of total sequenced reads to allow for easy reference to proportion of reads and also to account for the impact of sequencing depth.

2.5 Real-time PCR

Real-time PCR with SyBr Green was performed with 10 µl reactions using the FirePol MasterMix (Solis Biodyne) with 1 µl of 1:5 dilution of cDNA prepared previously for sequencing and 0.5 µl of 10 µM primers. For subgenomic RNA, we used covid_1 primer_left and covid_113/133/137/139/143/145/146/148 primer_right from the pool PCR primer set (Table 1) for qPCR.

2.6 Taqman assay

As there were subgenomic amplicons detected at low viral titres at Day 13 and 15 that reflected non-specific amplification with SyBr Green qPCR using ARTIC primers, we wanted to compare how one-step Taqman qRT-PCR assay fared compared to RT followed by SyBr Green qPCR. Taqman PCR were designed with one primer in the TRS-L and the other primer and the Taqman probe located within the subgenomic fragment about 20,000 nuclceotides away. Taqman probes resulted in better concordance between subgenomic RNA and genomic RNA and can be is empirically more sensitive compared to the two-step qRT-PCR as seen from the ability to detect the subgenomic RNA on P07 Day 12 (Supplementary Figure S3).

Real time probe qPCR measurements were performed Applied Biosystems QuantStudio system. Primers were designed to span 7b-mrna (seq_143) and E-mRNA (seq_133) respectively (Table 2) and the experiment was conducted as per TOYOBO THUNDERBIRD™ Probe qPCR Master Mix instructions. For SARS-COV-2 genomic RNA, 1 µl of heat-inactivated RNA was amplified using a 10 µl Fortitude 2.1 mix (MiRXES) as per manufacturer's instructions

TABLE 1 ARTIC primer design.

Primer name	Primer left	Primer right
covid_1	ACCAACCAACTTTCGATCTCTTGT	CTCGTTGAAACCAGGGACAAGG
covid_2	CAGCACATCTAGGTTTCGTCCG	CGAGCATCCGAACGTTTGATGA
covid_3	GGCACTTGTGGCTTAGTAGAAGT	AAATGACTTTAGATCGGCGCCG
covid_4	GGCGAAATACCAGTGGCTTACC	ATGCACTCAAGAGGGTAGCCAT
covid_5	TTAACGGAGGGGCATACACTCG	GGACATTCCCCATTGAAGGTGT
covid_6	AGAGCTATGAATTGCAGACACCTT	CGTCTGCCATGAAGTTTCACCA
covid_7	GCGTCACCAAATGAATGCAACC	ATAGTGCGACCACCCTTACGAA
covid_8	ACCTGAGCATAGTCTTGCCGAA	TGGCGATCTCTTCATTAAGTTTAAAGTCA
covid_9	TGACAACCTTCTTGAAATACTCCAAAAA	TACAACACGAGCAGCCTCTGAT
covid_10	GGTGCCTGGAATATTGGTGAACA	ACACCACCTGTAATGTAGGCCA
covid_11	ACTGAGACTCATTGATGCTATGATGT	ACAGGTGACAATTTGTCCACCG
covid_12	AGAGTTTCTTAGAGACGGTTGGGA	GCATGAGTAGGCCAGTTTCTTCT
covid_13	ATTTGTCACGCACTCAAAGGGA	TTTCGAGCAACATAAGCCCGTT
covid_14	ACAACCTACTAGTGAAGCTGTTGAA	ACCGAGTTCAACTGTATAGGCAGA
covid_15	TGAATATCACTTTTGAACTTGATGAAAGGA	ACCTTCTTCTTCATCCTCATCTGGA
covid_16	TGAGTCTGGTGAGTTTAAATTGGCT	GATTGTCCTCACTGCCGTCTTG
covid_17	TCAACCTGAAGAAGAGCAAGAAGAA	TGGCTGCATTAACAACCACTGT
covid_18	TGGAACTTACACCAGTTGTTCAGAC	TGTTTAGCAAGATTGTGTCCGCT
covid_19	TGATGATTACATAGCTACTAATGGACCAC	ACAGCTAAGTAGACATTTGTGCGA
covid_20	AGCTGGTATTTTTGGTGCTGACC	GTGAGGAACTTAGTTTCTTCCAGAGT
covid_21	CCTTCAGTTGAACAGAGAAAACAAGAT	CAGTAGTGCCACCAGCCTTTTT
covid_22	GGGTGATGTTGTTCAAGAGGGT	CATGTGCAAGCATTTCTCGCAA
covid_23	AAAAGTGCCTTTTACATTCTACCATCTATT	GTGTGTTGATAAGTGACGCTACAGT
covid_24	GGGTGTGGTTGATTATGGTGCT	ACCAGCAAGTGAGATGGTTTCA
covid_25	ACAGCGTATAATGGTTATCTTACTTCTTCT	CACAACTTGCGTGTGGAGGTTA
covid_26	CTTCTTTCTTTGAGAGAAGTGAGGACT	TGCTGACATGTACCTACCCAGA
covid_27	CACTCTACGTGTTGAGGCTTTTGA	GCACAAAAGTTAGCAGCTTCACC
covid_28	AGTTGAAGTTTAATCCACCTGCTCT	GTGTGCCCATGTACATAACAGCT
covid_29	ACGTGGTGTGTAAAACTTGTGGA	ACCGTCTATGCAATACAAAGTTTCTTT
covid_30	TGGTACATTTACTTGTGCTAGTGAGT	ACAAGATCAATTGGTTGCTCTGTGA
covid_31	GGTGTTGTTTGTACAGAAATTGACCC	AGCCACCACATCACCATTTAAGT
covid_32	AATTTGCTGATGATTTAAACCAGTTAACTG	CGCGTCCTCTGACTTCAGTACA
covid_33	GGTGTATACGTTGTCTTTGGAGCA	TGTGTGGCCAACCTCTTCTGTA
covid_34	ACTACCGAAGTTGTAGGAGACATTATACT	ACACCGTGTAACTATGTTAGTAGTTGT
covid_35	AGTGTCCCTTGGGATACTATAGCT	AAATGAAGCCTCTAGACAAAATTTACCG
covid_36	AATTCTAGAATTAAAGCATCTATGCCGAC	ACCAGTACAGTAGGTTGCAATAGTG
covid_37	GGCATGCCTTCTTACTGTACTGG	TGCAGCCAATCCAAGTACATAGAA
covid_38	TTTGGCTTAGTTGCAGAGTGGT	ACAACCGTCTACAACATGCACA
covid_39	CCCGATTTCAGCTATGGTTAGAATGT	ACTGTAGTGACAAGTCTCTCGCA

(Continued on following page)

TABLE 1 (*Continued*) ARTIC primer design.

Primer name	Primer left	Primer right
covid_40	TTGTGTTAATTGTGATACATTCTGTGCT	AGGCAATGAACCTTTAGTGTTATTAGCT
covid_41	AGCTGGTCAAAAGACTTATGAAAGACA	AACATTTTAACTGCAACTTCCGCAC
covid_42	ACAGTCAGCTTATGTGTCAACCTAT	CTGAATCAACAAACCCTTGCCG
covid_43	GCAGAAGCTGAACTTGCAAAGAA	TGACTTTTTGCTACCTGCGCAT
covid_44	TGAAAACATGACACCCCGTGAC	TTACCACCCTTAAGTGCTATCTTTGT
covid_45	GAATAACTTACCTTTTAAGTTGACATGTGC	TCACGAGTGACACCACCATCAA
covid_46	ACACCTGTTCATGTCATGTCTAAACA	TTGTGCGTAATATCGTGCCAGG
covid_47	GCCCATTGATTGCTGCAGTCAT	CGTGTGTCAGGGCGTAAACTTT
covid_48	TGGTAAGCCAGTACCATATTGTTATGA	TCTACACCACAGAAAACTCCTGGT
covid_49	GCTGGTGTTTGTGTATCTACTAGTGG	GGCAACTACATGACTGTATTCACCA
covid_50	TCGTAGTAACATGCCTTGCCTAC	CCAGAAAGGTACTAAAGGTGTGAACA
covid_51	CCAGTTTACTCATTCTTACCTGGTGT	TTAACAAAAAGGTGCACAGCGC
covid_52	TGGTTCTTTAGTAATTACCTAAAGAGACGT	TGAGAGCCTTTGCGAGATGACA
covid_53	AGTACAAGTATTTTAGTGGAGCAATGGA	AGCATGTCTTCAGAGGTGCAGA
covid_54	ACACTTAACGGTCTTTGGCTTGA	CTGTCCTGGTTGAATGCGAACA
covid_55	TGTACTTAAGCTTAAGGTTGATACAGCC	ACCTTCTAAGTCTGTGCCAGCA
covid_56	ACTGTGTCTCTTTTTGTTACATGCAC	GTCAACATGGTCTTGTGTTAGAGGT
covid_57	GTTTCTCAATCGATTTACCACAACTCT	AACCAGTGGTGTGTACCCTTGA
covid_58	TGTTGTTAGACAATGCTCAGGTGT	AGCTACAGTGGCAAGAGAAGGT
covid_59	TGCTATGGGTATTATTGCTATGTCTGC	GTCCACACTCTCCTAGCACCAT
covid_60	ACTGTGTTATGTATGCATCAGCTGT	AGAAAATAGGGCAATACTCAACACACA
covid_61	TCTCTGTTACTTCTAACTACTCAGGTGT	TGCTATTCTTGGGTGGGAGTAGT
covid_62	ACTGACTCTTGGTGTTTATGATTACTTAGT	CTGGACACATTGAGCCCACAAT
covid_63	TGCACATCAGTAGTCTTACTCTCAGT	AGCTGCATATGATGGAAGGGAAC
covid_64	AGCTTTGTGAAGAAATGCTGGACA	TTGCCCTCTTGTCCTCAGATCT
covid_65	CGTAAGTTGGAAAAGATGGCTGATC	GGATTTCCCACAATGCTGATGC
covid_66	TGGTTGTCATACCAGACTATAACACAT	TAGTACCGGCAGCACAAGACAT
covid_67	GCTTTAAGGGCCAATTCTGCTG	CCTACAAGGTGGTTCCAGTTCTG
covid_68	GGATTTGAAATGGGCTAGATTCCCT	TTGGTTGTCCCCCACTAGCTAG
covid_69	GCCAATTCAACTGTATTATCTTTCTGTGC	CCCACAGGGTCATTAGCACAAG
covid_70	CTGTACTGCCGTTGCCACATAG	GTAAGACGGGCTGCACTTACAC
covid_71	GAACCCATGCTTCAGTCAGCTG	GCAACAGCTGGACAATCCTTAAGT
covid_72	TACTTTGTAGTTAAGAGACACACTTTCTCT	GGGTTTTCTACAAAATCATACCAGTCCT
covid_73	TGTGACACATTAAAAGAAATACTTGTCACA	ACCTGGCGTGGTTTGTATGAAA
covid_74	TGGTATTGTTGGTGTACTGACATTAGA	TGGGTGGTATGTCTGATCCCAA
covid_75	TGGGATTTGTTAAAATATGACTTCACGG	ACACCTAGCTCTCTGAAGTGGT
covid_76	CAAGTTTTGGACCACTAGTGAGAAAAA	CCGGGTTTGACAGTTTGAAAAGC
covid_77	ATGCACGCTGCTTCTGGTAATC	AGTTGTCTGATATCACACATTGTTGGT
covid_78	TCTTTGCTCAGGATGGTAATGCT	TGCGAAAAGTGCATCTTGATCCT

(Continued on following page)

TABLE 1 (*Continued*) ARTIC primer design.

Primer name	Primer left	Primer right
covid_79	CAACAACCTAGACAAATCAGCTGGT	CTCCTCTAGTGGCGGCTATTGA
covid_80	CCGTAGCTGGTGTCTCTATCTGT	ACACGTTGTATGTTTGCGAGCA
covid_81	CCTAAATGTGATAGAGCCATGCCT	TGCATTAACATTGGCCGTGACA
covid_82	CATCAGGAGATGCCACAACTGC	AGCCACTAGACCTTGAGATGCA
covid_83	ACATTTCTCAATGATGATACTCTCTGACG	GGCCCCTAGGATTCTTGATGGA
covid_84	ACATACAATGCTAGTTAAACAGGGTGA	ACATGTGTCCTGTTAACTCATCATGT
covid_85	ATGCTTACCCACTTACTAAACATCCT	GGTCTACGTATGCAAGCACCAC
covid_86	ACCGCATACAGTCTTACAGGCT	AGCACACAATGGAAAACTAATGGGT
covid_87	GATGTGACTCAACTTTACTTAGGAGGT	ACAGCACTTCACGTACAGTAGC
covid_88	AAGCTTTTTGCAGCAGAAACGC	GTTGTACCTCGGTAAACAACAGCA
covid_89	TGGTTATCGTGTAACTAAAAACAGTAAAGT	GGTGGTCCCTGGAGTGTAGAAT
covid_90	CTCAGATGAGTTTTCTAGCAATGTTGC	TCAAAACACTCTACACGAGCACG
covid_91	GATGCACTATGTGAGAAGGCATTAAAA	GTGCAGGTAATTGAGCAGGGTC
covid_92	TGATTTGAGTGTTGTCAATGCCAGA	ACATCATGCGTGATAACACCCT
covid_93	CACTGTGAGTGCTTTGGTTTATGAT	GAGCCCTGTGATGAATCAACAGT
covid_94	TGGAGAAAAGCTGTCTTTATTTCACCT	TGTAAAGTTGCCACATTCCTACGT
covid_95	TTACCAGAGCAAAAGTAGGCATACT	TCATGTCCTTAGGTATGCCAGGT
covid_96	TACACAGGCACCTACACACCTC	AGGTACAGCAACTAGGTTAACACC
covid_97	GAGGGGTGTCATGCTACTAGAGA	TGTCAACTCAAAGCCATGTGCC
covid_98	CCTTGGAATGTAGTGCGTATAAAGATTG	TGGTTGCTTTGTAGGTTACCTGT
covid_99	ATTGGATTTGATTACGTCTATAATCCGTTT	CCAATGTCGTGAAGAACTGGGA
covid_100	ATGCGGCTTGTAGAAAGGTTCA	CGACATTGCAATTCCAAAATAGGCA
covid_101	CCTTGTAGTGACAAAGCTTATAAAATAGAAG	ACATGGACTGTCAGAGTAATAGAAAAATG
covid_102	CATTCCACACACCAGCTTTTGATAA	AAACCCACAAGCTAAAGCCAGC
covid_103	AGACATCATGCTAATGAGTACAGATTGT	GCCCAAAGCTCAAATGCTACATT
covid_104	ACACAAAAGTTGATGGTGTTGATGT	CAGTGAGTGGTGCACAAATCGT
covid_105	GCTCCAGCACATATATCTACTATTGGT	ACTGTGTTTTTACGGCTTCTCCA
covid_106	ACAACCATCTGTAGGTCCCAAAC	TCAGTAGATGTAAACCACCTAACTGACT
covid_107	GGTATAAATTAGAAGGCTATGCCTTCG	CTTTGACAACCTTAGAAACTACAGATAAATC
covid_108	CGCAAACAGGTTCATCTAAGTGTG	GCCTTTAGGTAATGTTGCACTATCAC
covid_109	TGCTATGCCTAATCTTTACAAAATGCA	GATCTGAATCGACAAGCAGCGT
covid_110	AAAGGAGTTGCACCAGGTACAG	ATAGCCACGGAACCTCCAAGAG
covid_111	TGTTACAAAAGAAAATGACTCTAAAGAGGG	TAACCATCTATTTGTTCGCGTGGT
covid_112	AATGTGAATGCGTCATCATCTGAAG	GAACATCACTAGAAATAACAACTCTGTTGT
covid_113	ACTGCTGTTATGTCTTTAAAAGAAGGTCA	CAGGGTAATAAACACCACGTGTGA
covid_114	AGTCAGTGTGTTAATCTTACAACCAGA	TGTTAGACTTCTCAGTGGAAGCAAA
covid_115	TGGGACCAATGGTACTAAGAGGT	ACTCTGAACTCACTTTCCATCCAAC
covid_116	CGCTACTAATGTTGTTATTAAAGTCTGTGA	CCTGAGGGAGATCACGCACTAA
covid_117	AGGGAATTTGTGTTTAAGAATATTGATGGT	GCACAGTCTACAGCATCTGTAATGG

(Continued on following page)

10.3389/fgene.2023.1086865

TABLE 1 (*Continued*) ARTIC primer design.

Primer name	Primer left	Primer right
covid_118	TGTGGGTTATCTTCAACCTAGGACT	AACAGATGCAAATCTGGTGGCG
covid_119	AGAGTCCAACCAACAGAATCTATTGT	CCCTGGAGCGATTTGTCTGACT
covid_120	TGATCTCTGCTTTACTAATGTCTATGCA	TGTGCTACCGGCCTGATAGATT
covid_121	AACAATCTTGATTCTAAGGTTGGTGGT	TAGGTCCACAAACAGTTGCTGG
covid_122	GGTTTCCAACCCACTAATGGTGT	TCAAGTGTCTGTGGATCACGGA
covid_123	ACAAAAAGTTTCTGCCTTTCCAACA	CCCCTATTAAACAGCCTGCACG
covid_124	CAACTTACTCCTACTTGGCGTGT	AAAATTTGTGGGTATGGCAATAGAGTT
covid_125	TGCCTACACTATGTCACTTGGTG	CTTGTGCAAAAACTTCTTGGGTGT
covid_126	TGGCAGTTTTTGTACACAATTAAACCG	AGCAGCAATATCACCAAGGCAA
covid_127	ACTTTTCAACAAAGTGACACTTGCA	AGAACATTCTGTGTAACTCCAATACCA
covid_128	CAGGTGCTGCATTACAAATACCATT	GCCTCAACTTTGTCAAGACGTGA
covid_129	ACGCTTGTTAAACAACTTAGCTCCA	ATGAGGTGCTGACTGAGGGAAG
covid_130	GTCAGAGTGTGTACTTGGACAATCA	ACAACATCACAGTTACCAGACACA
covid_131	TGGCACACACTGGTTTGTAACA	AGGCGGTCAATTTCTTTTTGAATGT
covid_132	TCACCAGATGTTGATTTAGGTGACA	GCAGCAGGATCCACAAGAACAA
covid_133	TGGTGACAATTATGCTTTGCTGTATG	GCAACGCCAACAATAAGCCATC
covid_134	CTCCTTCAGATTTTGTTCGCGC	GATAGAGAAAAGGGGCTTCAAGGC
covid_135	GGGTGTTCACTTTGTTTGCAACT	GACTTGTTGTGCCATCACCTGA
covid_136	GCCAACTATTTTCTTTGCTGGCA	ACATGTTCAACACCAGTGTCTGT
covid_137	TGTATTACACAGTTACTTCACTTCAGACT	CGTACCTGTCTCTTCCGAAACG
covid_138	GACGACTACTAGCGTGCCTTTG	TCGTTTAGACCAGAAGATCAGGAACT
covid_139	ACGTGAGTCTTGTAAAACCTTCTTTTT	TGGCATAGGCAAATTGTAGAAGACA
covid_140	AAGCTCCTTGAACAATGGAACCT	AATGACCACATGGAACGCGTAC
covid_141	TAGGCTTGATGTGGCTCAGCTA	AGCGTTCGTGATGTAGCAACAG
covid_142	TTGCTGGACACCATCTAGGACG	ACCTGAAAGTCAACGAGATGAAACA
covid_143	ACACAGACCATTCCAGTAGCAGT	AGCGAGTGTTATCAGTGCCAAG
covid_144	TGAAGAGCAACCAATGGAGATTGA	GTGTTTTACGCCGTCAGGACAA
covid_145	TCCTCTAGCTGATAACAAATTTGCACT	AGCAGAAAGGCTAAAAAGCACAAA
covid_146	TGCTTCACACTCAAAAGAAAGACAGA	AGGACACGGGTCATCAACTACA
covid_147	ACAACTGTAGCTGCATTTCACCA	ACGAACAACGCACTACAAGACT
covid_148	ACCCATTCAGTACATCGATATCGGT	ACTGCCAGTTGAATCTGAGGGT
covid_149	TAATGGACCCCAAAATCAGCGA	CGTCTGGTAGCTCTTCGGTAGT
covid_150	CGAGGACAAGGCGTTCCAATTA	CGATTGCAGCATTGTTAGCAGG
covid_151	TGCAACTGAGGGAGCCTTGAAT	TCAATCTGTCAAGCAGCAGCAA
covid_152	GGAACTTCTCCTGCTAGAATGGC	GTCTGATTAGTTCCTGGTCCCCA
covid_153	TAACACAAGCTTTCGGCAGACG	TAGGCTCTGTTGGTGGGAATGT
covid_154	TCCAAATTTCAAAGATCAAGTCATTTTGC	CCTTGTGTGGTCTGCATGAGTT
covid_155	TTCTCCAAACAATTGCAACAATCCA	CGGTGAAAATGTGGTGGCTCTT
covid_156	ACTCTTGTGCAGAATGAATTCTCGT	AAGCTATTAAAATCACATGGGGATAGCA

TABLE 2 Primers and Taqman probes for Taqman assay.

Name	Sequence
seq_133_reverse	GCCATAACAGCCAGAGGAAA
seq_143_reverse	GCCGTCAGGACAAGCAAAAG
seq_left	ACCTTCCCAGGTAACAAACCA
seq_133_internal_probe	[6FAM]AATTTGCCTATGCCAACAGG[BHQ1]
seq_143_internal_probe	[6FAM]TTGGCACTGATAACACTCGC[BHQ1]

(scaled to 10 µl reaction). For subgenomic RNA, 1 µl of heat-inactivated RNA was mixed with 5 µl qPCR mastermix, 0.25 µl DNA polymerase, 0.25µl RT, 0.25 µl 10 µM probe, 0.5 µl 10 µM primers in a 10 µl reaction. For both reactions, the cycling conditions were as followed, 48°C 15 min, 95°C 2 min, and 45 cycles of 95°C 10s, 59°C 30s.

2.7 Real-time PCR analysis

Quantification of the relative abundance of sub-genomic transcripts to genomic transcripts were performed by obtaining delta Ct change defined by the difference between the geometric mean of the Taqman assay targeting sub-genomic transcripts described above and the fortitude assay. Wilcoxon ranked sum test with continuity correction is then applied to calculate the significance of the delta CT changes across time.

3 Results

3.1 Amplicon-based sequencing identifies sub-genomic transcript-derived junction reads

In order to probe genomic changes in SARS-CoV-2, we designed 156 primers pairs using the ARTIC protocol (https://artic.network) to generate genome tiling amplicons approximately 300 nt in length to be compatible with 2 × 150bp sequencing chemistry. This protocol allows generation of consensus COVID-19 genome and dentification of single nucleotide variants (Figure 1A). We sequenced samples from 53 SARS-CoV-2 positive patients who were hospitalized at the National University Health System between February to May 2020. We observed that a proportion of sequencing reads does not contain the intended paired primer sequences. These reads span large genomic distances, indicative of gapped reads originating from the splice junction of the subgenomic transcripts (Kim et al., 2020). Alignment of the gap-spanning junction reads to canonical subgenomic transcripts (Figure 1B) identified all nine previously reported canonical subgenomic transcripts (Wölfel et al., 2020), indicating that our amplicon design generated reads composed of leader-body junctions. The genome distribution of junction reads shows a sharp peak at the 5' end corresponding to the leader sequence and high coverage in the 3' end for the nested transcript bodies (Figure 1B). These canonical gapped reads made up 0.001%–1% of the total amplicon reads. Estimating the transcript abundance using junction reads across all 53 samples showed varying levels of total subgenomic transcripts across all samples and transcription distribution within each sample (Figures 2A,B). Notably, we see in some samples, a high expression of the envelope encoding gene (M)

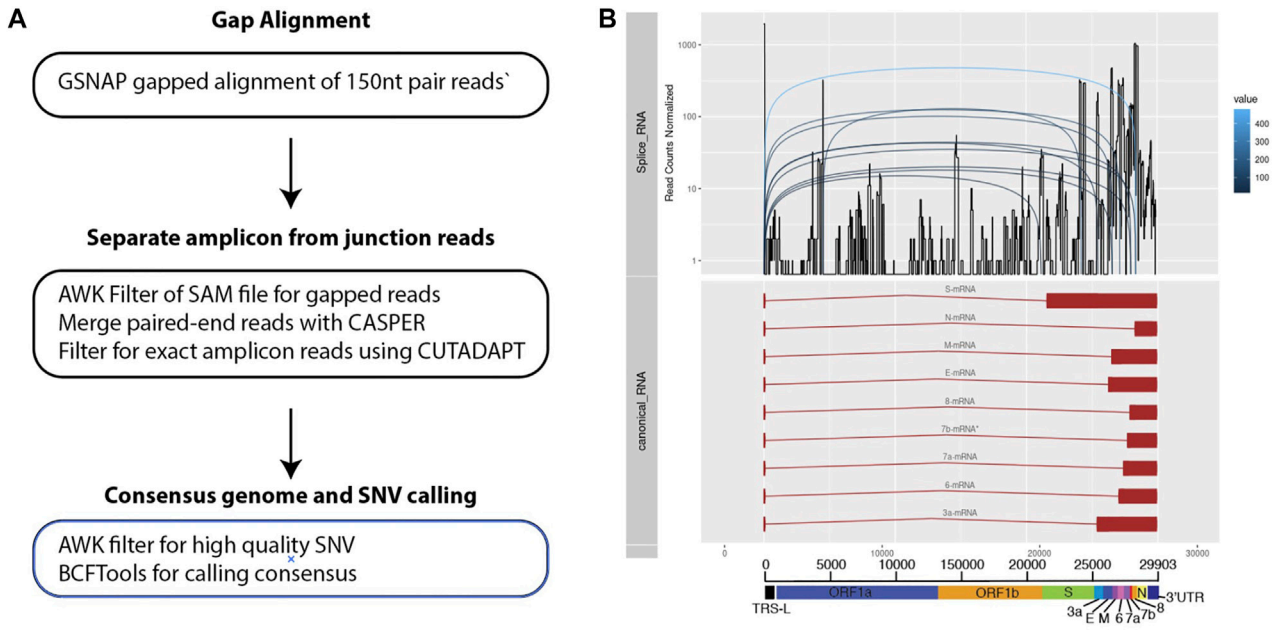

FIGURE 1
Bioinformatics workflow for identification and visualization of gapped reads and amplicons. **(A)** Three main stages of processing the sequenced data. **(B)** Multiple track visualization of the sequenced data. X-axis corresponds to the COVID genomic length. The top track of the figure reflects the structure of the known and annotated sequences of subgenomic transcripts. The second tracks show an example of the coverage of junctions reads identified using the pipeline specified. Distinct peaks are observed that corresponds to the canonical transcripts on the top track. The third tracks reflect how gapped reads are separated across the subgenomic junctions.

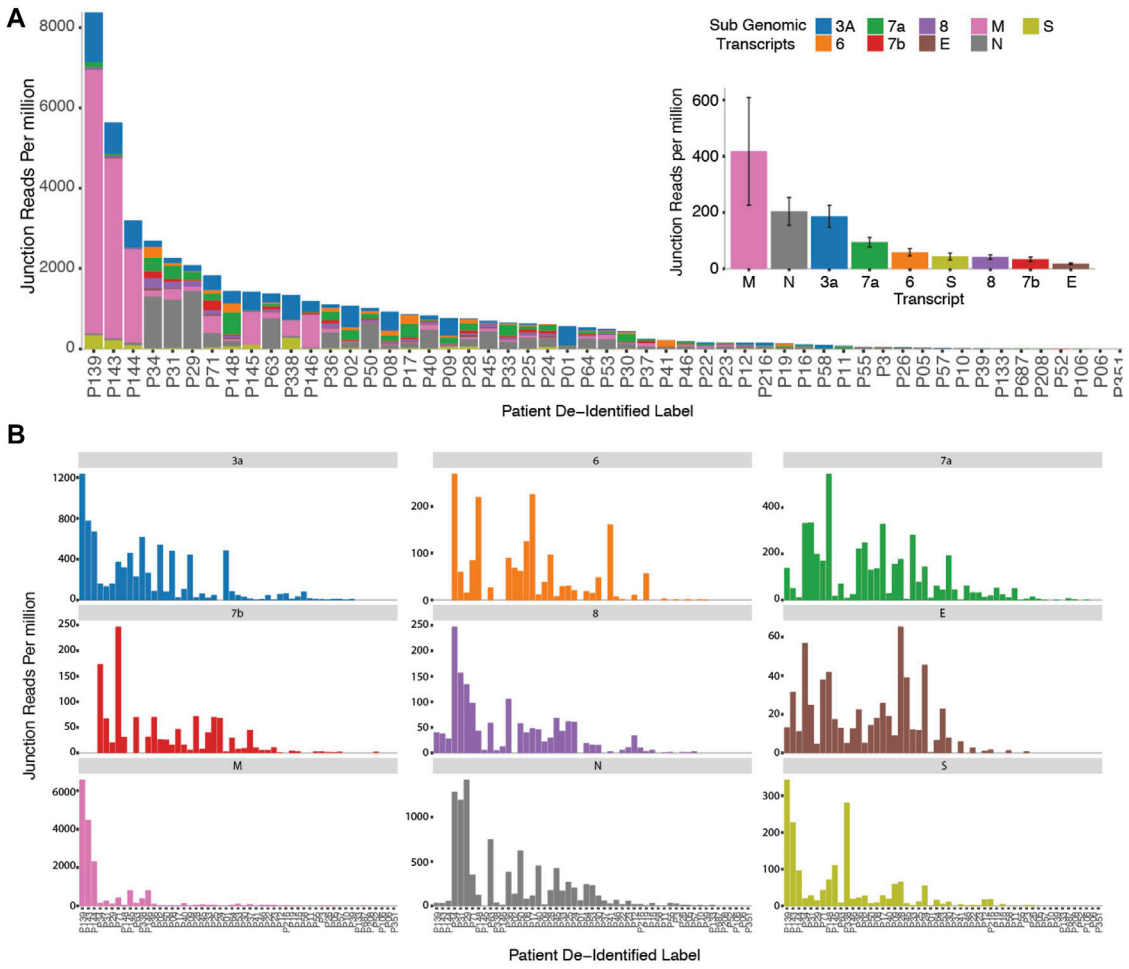

FIGURE 2
Cohort wide identification of subgenomic Reads. **(A)** Stacked bar graphs reflecting the amount of normalized sub genomic gapped reads for each of the nine canonical subgenomic transcripts across our patient cohort. (Inset) Distribution of total reads per canonical transcript across the patient cohort. **(B)** Distribution of normalized individual subgenomic gapped read counts across the patient cohorts.

distinct from that seen in transcripts isolated from cultured virus (Kim et al., 2020).

3.2 Tracking abundance of subgenomic transcripts across infection course

Next, we investigated the effect of viral titre on genome and subgenomic read recovery. As the ORF1ab region targeted by Fortitude 2.1 is not included in any canonical subgenomic RNA, it is actually a good proxy for genomic RNA. We found that the degree of genome read coverage was lower at higher Ct (Figure 3A). This observation is consistent with the hypothesis that high Ct samples are dominated by viral fragments rather than intact viruses (Hu et al., 2020). Comparing genome coverage against the day of illness demonstrated that as the viral load decreased over the course of the infection, the genome coverage generated by our protocol correspondingly decreases (Figure 3B). This suggests an explanation why subgenomic reads were not recovered in samples that were of higher CT values. To explore this more quantitatively, we performed qPCR with Taqman assays (Fortitude 2.1 and custom-designed subgenomic assays targeting ORF5 and ORF7a subgenomic RNA) that specifically target junctional subgenomic transcripts and non-junctional genomic RNA across the 53 patient samples (Figure 3C). We confirmed a general trend where the CT values of both junctional and non-junctional targets increased over the course of the illness. We further observed a small but significant increase in junctional targets CT compared to non-junctional targets in the latter stages of the infection (Figure 3D), suggesting that subgenomic transcripts may possibly serve as an early measure of infection resolution although the variability observed in subgenomic across samples grouped by day of illness was high.

3.3 Subgenomic junctional reads captures temporal variation

This variability in subgenomic titres across samples may reflect differences in sample quality, processing or patient-to-patient variation in treatment protocol or intra-host variation or differences in disease trajectory. To better evaluate the utility of subgenomic quantitation in tracking disease progression, we

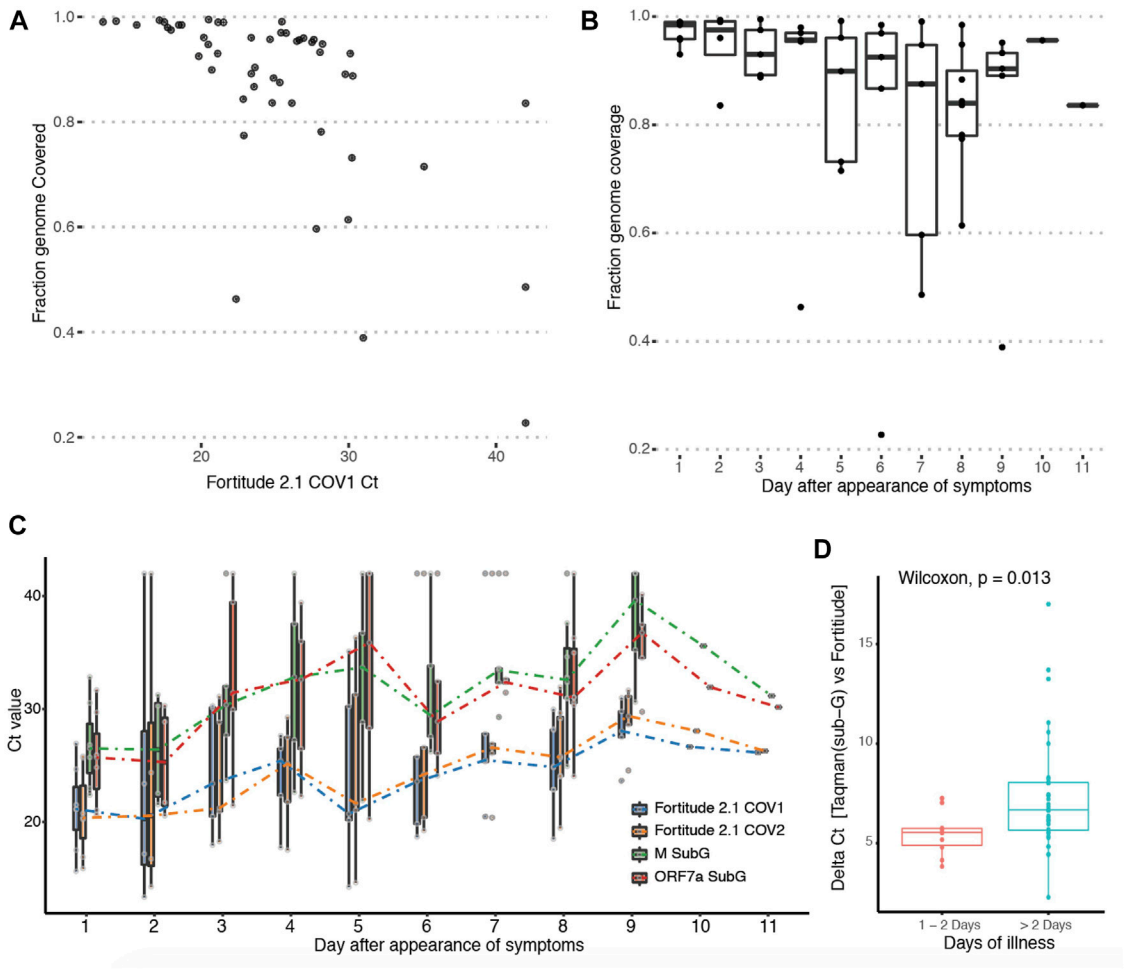

FIGURE 3
Tracking abundance of subgenomic transcripts across infection course. **(A)** Y-axis: Genome coverage by Arctic sequencing across 53 patients. X-axis: Ct-value of the Taqman assay targeting the genomic region of the SARS-Cov-2 genome from each patient. **(B)** Boxplot of genomic coverage by Arctic sequencing binned by day of symptom at which patient was sampled. **(C)** Boxplot of Taqman quantitation Ct values using primers targeting genomic-specific RNA (Fortitude 2.1 COV1 and COV2) and subgenomic-specific RNA (3a SubG, 7a SubG) binned by day of symptom at which patient was sampled. **(D)** Wilcoxon rank sum test with continuity correction was performed on samples binned by early (day 1 and 2) and latter stages (day 3–11) of infection.

sampled a single patient (P07) daily from day 6 of illness to day 15 (Figure 4A). At each time point, the collected samples were sequenced. Total and subgenomic junctional reads were quantified. The patient was administered lopinavir/ritonavir daily from day 10. We observed that coverage across the genome dropped significantly by more than 50% on day 11 (Figure 4B), the day after antiviral treatment was administered. Notably, subgenomic junction reads across all canonical transcripts also reduced significantly on day 9 and was no longer detectable by day 11 (Figure 4C). Individual subgenomic transcripts showed a similar trend. To verify these, we used amplicon primers that target genomic RNA (dotted line; targeted to N gene region) and junction spanning qPCR (Figure 5A) to quantify genomic and subgenomic RNA targets respectively. For convenience of plotting, undetermined Ct values by qPCR were assigned the value of 42. The amount of subgenomic RNA detected by qPCR tracked the viral genome signal, but fell drastically at day 9 or 10 to undetectable levels on day 11 in agreement with the sequencing data. The amplicon at the completion of the qPCR was then analyzed by agarose gel electrophoresis to verify that amplicons were of the expected sizes (Figure 5B). Importantly, a synthetic SARS-CoV-2 RNA control (Twist) was used as a negative PCR control to demonstrate that the PCR primers were specific for gap-spanning templates as the Twist control template is composed only of synthetic SARS-CoV-2 genome provided as 5 kb fragments. We repeated the analysis with another different longitudinal patient sample (P08) and showed that subgenomic RNA levels tracked genomic RNA similarly, with an earlier decline in the progression of disease (Supplementary Figure S2).

4 Discussion

In this article, we present a method for workflow and Taqman probes for the detection of subgenomic RNA which may have diagnostic value in the COVID-19 pandemic, potentially for determining the infectivity of patients. However, further studies will need to be done to validate this hypothesis. Other than

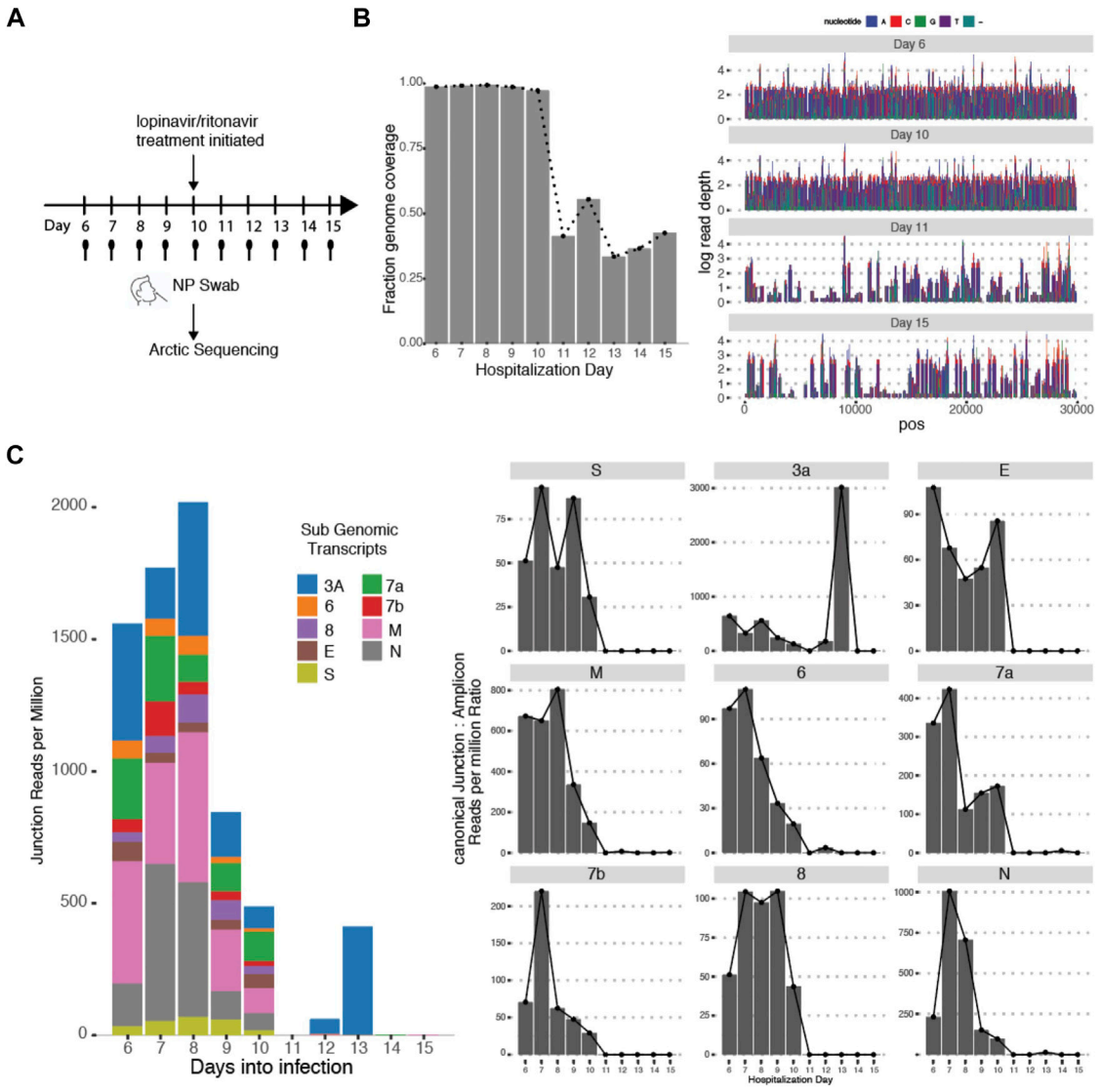

FIGURE 4
Subgenomic Junctional Reads captures temporal variation. **(A)** Sampling protocol of Patient P07. Patient was sampled daily from day 6 of hospitalization. Patient was treated with Lopinavir/ritonavir from day 10 onwards. **(B)** (left) Genome coverage by Arctic sequencing across the 10 timepoints. (right) Raw read coverage across the genome for Day 6, 10,11,15. Sequencing depth was similar across all timepoints (Supplementary Figure S1) **(C)** (left) Stacked bar graphs reflecting the amount of normalized sub genomic gapped reads for each of the nine canonical subgenomic transcripts across the 10 timepoints. (right) Distribution of normalized individual subgenomic gapped read counts across the 10 timepoints.

ORF1ab, which spans roughly two-thirds of the coronavirus genome and codes for replicase-transcriptase proteins, the other polypeptides including the spike and envelope proteins are dependent on the production of negative strand subgenomic RNA that has the transcription-regulating sequence (TRS) "spliced" next to the translational start site of these polypeptide (Sawicki et al., 2007). This allows transcription of 50–100-fold excess of viral mRNAs that ultimately produce proteins essential for virion formation. Disrupting the formation of N gene subgenomic RNA and consequently N gene mRNA through mutation of conserved RNA motifs effectively curtailed the ability of another coronavirus—the transmissible gastroenteritis virus to produce successful virions (Mateos-Gomez et al., 2013).

Absence of subgenomic RNA in infected cells will likely mean the infected cells are not capable of producing active virions. If active infectious virions are produced and actively transmitting between cells within the host, we hypothesize there will be subgenomic RNA produced. As production of subgenomic RNA is dependent on the intact replicase/transcriptase complex skipping a large portion of the genome and not produced spontaneously from just reverse transcription of the virus, it is distinct from viral fragment residues present in the virus at late stages of infection (Hu et al., 2020) and may be more valuable in detection of infectious SARS-COV-2 patients. Thus, it may be possible to use subgenomic RNA levels to ascertain the presence of 'active' infectious virions in which the virions can infect the naïve host cells at the site of sampling. Indeed, when we correlate our subgenomic RNA levels or proportion with day of symptom, we see a correlation that is consistent with the predicted transmissibility of SARS-CoV-2 (He et al., 2020), suggesting that this hypothesis is worth

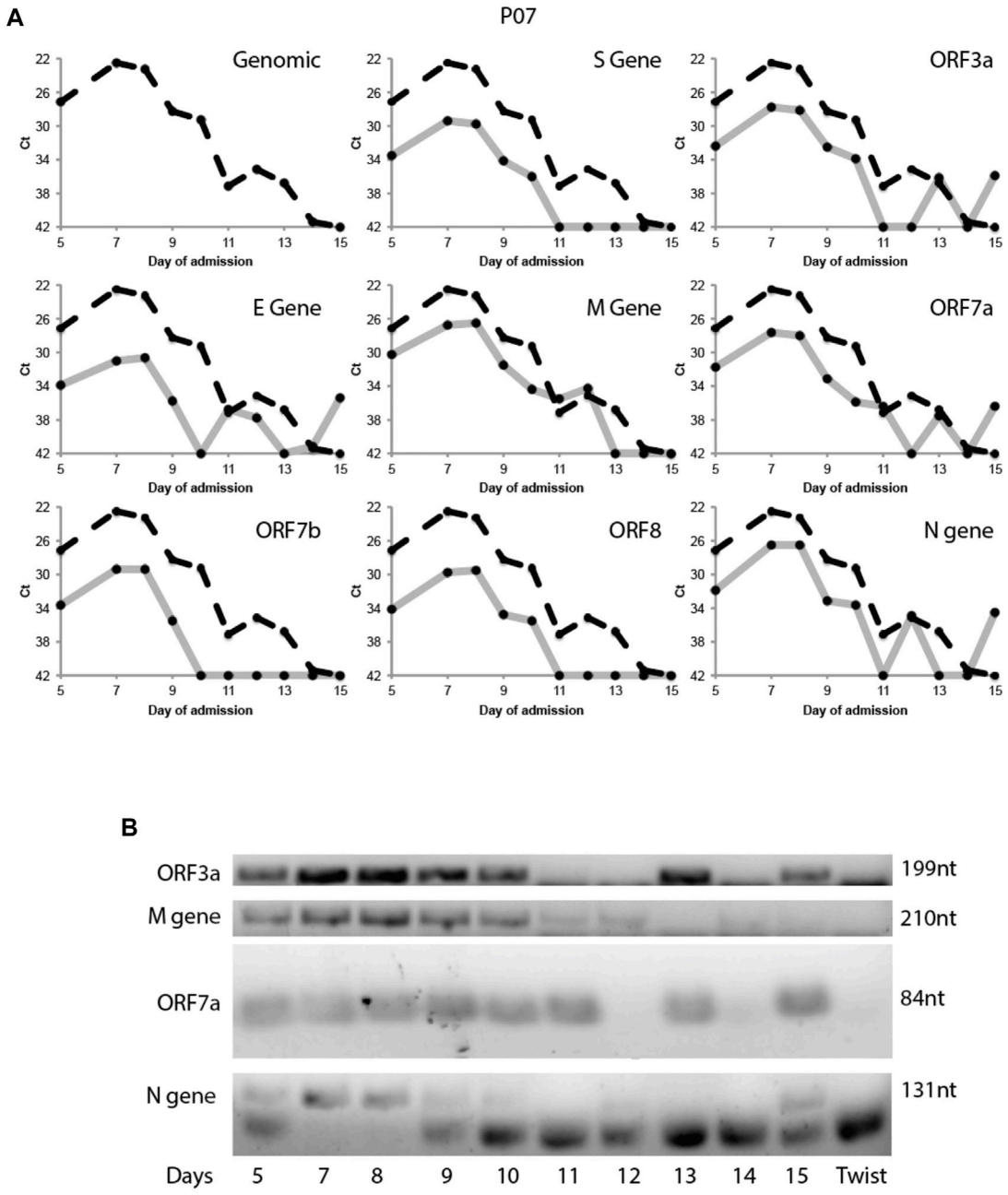

FIGURE 5
Tracking Patient 07 subgenomic expression from nasopharyngeal swabs across days of admission. Taqman probes were used to amplify a genomic target and eight subgenomic RNA regions specifically target the discontinuous junction *via* real time RT-PCR. **(A)** The Ct values of the subgenomic RNAs (solid grey line) were plotted across 10 time points together with the genomic Ct (dotted black line). **(B)** The PCR product of the above qRT-PCR were verified by gel electrophoresis. Synthetic SARS-CoV-2 genomic RNA (TWIST Bioscience) was used as a negative control.

investigating further. There are a few caveats to this hypothesis. Firstly, subgenomic RNAs can self-amplify without the presence of a full-length genomic template, which means the positive strand can act as a template for the negative strand and *vice versa* (Wu and Brian, 2010). However, this requires the presence of ORF1ab proteins that are required for the generation of the subgenomic RNA in the first place, thus subgenomic RNAs are unlikely to persist long after degradation of the genomic template encoding ORF1ab. Second, a similar study on SARS-CoV-2 has suggested that subgenomic RNA content was poorly correlated to day of illness in two patients and thus not a universally useful indicator of viral replication (Alexandersen et al., 2020). They noted the observation that subgenomics samples were more abundantly detected in poor samples enriched in degraded RNA because of a bias for short reads during PCR amplification. This was based on a limited number of samples (n = 2) collected under different conditions and was not conclusive. We performed a controlled time-course experiment where samples were collected from the same patients over the course of 10 days and both sequencing and Taqman quantitation were performed. In this way, differences in

subgenomic RNA levels due to sample handling were minimized. We found that subgenomic RNA tracked the genomic RNA over the course of the infection and showed an earlier drop-off that coincided with antiviral treatment.

Lopinavir has shown *in vitro* activity against SARS-CoV-2 replication (Choy et al., 2020), and there was a noticeable reduction in SARS-CoV-2 genome coverage in our patient (P07) following lopinavir/ritonavir initiation, although the subgenomic RNA levels were already in decline a day prior to that. The decline in the viral genome coverage may be due to the activity of the drug, or it may be due to the natural viral kinetics during the disease course. The combination lopinavir/ritonavir is currently not recommended as a treatment regimen for patients with COVID-19 in view of a lack of significant therapeutic effect (WHO, 2023; COVID-19 Treatment Guidelines, 2021). Moving forward, it would be worth using subgenomic RNA levels as a marker for antiviral efficacy in this and other widely used therapeutics in COVID-19 patients such as remdesivir or dexamethasone.

In summary, we demonstrated the ability to detect subgenomic RNA using the ARTIC protocol with low sequence depth and showed that we can track subgenomic distribution and contribution against genomic signal in multiple patients. We showed that there is correlation between subgenomic percentage, subgenomic Ct, genome coverage and genomic Ct with time after symptom first appears. Importantly, we developed a Taqman assay against ORF5 and ORF7a subgenomic RNA that we believe can be used to more closely monitor if subgenomic RNA is reflective of infectivity and can readily be incorporated into other available SARS-COV-2 diagnostics. We believe that this will make a significant contribution to public health efforts to control the virus as well as therapeutic studies evaluating the effects of antiviral therapies.

Data availability statement

The data represented in the study are deposited in the figshare repository, with the accession number: 22060181 and associated doi link: https://doi.org/10.6084/m9.figshare.22060181.v1.

Ethics statement

The studies involving human participants were reviewed and approved by National University Hospital Singapore DSRB 2020/00867. The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The authors would like to acknowledge the help and contributions of members of the Molecular Engineering Laboratory for helpful discussions. This work is supported by Agency for Science, Technology and Research (A*STAR) and Industrial Alignment Fund Pre-Positioning (IAF-PP) grant H17/01/a0/012. All procedures performed in studies involving human participants were in accordance with the ethical standards of National University Health System and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study is covered under DSRB 2020/00867.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1086865/full#supplementary-material

References

Alexandersen, S., Chamings, A., and Bhatta, T. R. (2020). SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. *Nat. Commun.* 11 (1), 6059. doi:10.1038/s41467-020-19883-7

Bioconductor (2022). Bioconductor - rsamtools. Available at: https://bioconductor.org/packages/release/bioc/html/Rsamtools.html.

Choy, K. T., Wong, A. Y. L., Kaewpreedee, P., Sia, S. F., Chen, D., Hui, K. P. Y., et al. (2020). Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication *in vitro*. *Antivir. Res.* 178, 104786. doi:10.1016/j.antiviral.2020.104786

COVID-19 Treatment Guidelines (2021). Lopinavir/ritonavir and other HIV protease inhibitors | COVID-19 treatment Guidelines. Available at: https://www.covid19treatmentguidelines.nih.gov/therapies/antiviral-therapy/lopinavir-ritonavir-and-other-hiv-protease-inhibitors/.

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* 26 (5), 672–675. doi:10.1038/s41591-020-0869-5

Hu, F., Chen, F., Wang, Y., Xu, T., Tang, X., and Li, F. (2020). *Failed detection of the full-length genome of SARS-CoV-2 by ultra-deep sequencing from the recovered and discharged patients retested viral PCR positive*. medRxiv. 2020.03.27.20043299.

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell* 181 (4), 914–921. doi:10.1016/j.cell.2020.04.011

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinforma. [Internet]* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Mateos-Gomez, P. A., Morales, L., Zuñiga, S., Enjuanes, L., and Sola, I. (2013). Long-distance RNA-RNA interactions in the coronavirus genome form high-order structures promoting discontinuous RNA synthesis during transcription. *J. Virol.* 87 (1), 177–186. doi:10.1128/JVI.01782-12

Maurier, F., Beury, D., Fléchon, L., Varré, J. S., Touzet, H., Goffard, A., et al. (2019). A complete protocol for whole-genome sequencing of virus from clinical samples: Application to coronavirus OC43. *Virology* 531, 141–148. doi:10.1016/j.virol.2019.03.006

MIN COW (2020). *Position statement from the national centre for infectious diseases and the chapter of infectious disease physicians, academy of medicine, Singapore: Period of infectivity to inform strategies for de-isolation for COVID-19 patients*. Singapore: Elements.

Ministry of Health (2022). Revised discharge for covid-19 patients [internet]. Available at: https://www.moh.gov.sg/news-highlights/details/revised-discharge-criteria-for-covid-19-patients (Accessed Oct 26, 2022).

Oran, D. P., and Topol, E. J. (2020). Prevalence of asymptomatic SARS-CoV-2 infection: A narrative Review. *Ann. Intern Med.* 173 (5), 362–367. doi:10.7326/M20-3012

Sawicki, S. G., Sawicki, D. L., and Siddell, S. G. (2007). A contemporary view of coronavirus transcription. *J. Virol.* 81 (1), 20–29. doi:10.1128/JVI.01358-06

Sola, I., Almazán, F., Zúñiga, S., and Enjuanes, L. (2015). Continuous and discontinuous RNA synthesis in coronaviruses. *Annu. Rev. Virol.* 2 (1), 265–288. doi:10.1146/annurev-virology-100114-055218

WHO (2023). Solidarity clinical trial for COVID-19 treatments. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments.

Wickham, H. (2009). *ggplot2*. New York, NY. - References - Scientific Research Publishing: Springer. Available at: https://www.scirp.org/reference/ReferencesPapers.aspx?ReferenceID=2175727.

Wölfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M. A., et al. (2020). Virological assessment of hospitalized patients with COVID-2019. *Nature* 581 (7809), 465–469. doi:10.1038/s41586-020-2196-x

Wu, H. Y., and Brian, D. A. (2010). Subgenomic messenger RNA amplification in coronaviruses. *Proc. Natl. Acad. Sci. U. S. A.* 107 (27), 12257–12262. doi:10.1073/pnas.1000378107

Wu, T. D., and Watanabe, C. K. (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21 (9), 1859–1875. doi:10.1093/bioinformatics/bti310