Nanopore target sequencing for accurate and comprehensive detection of SARS-CoV-2 and other respiratory viruses

Ming Wang[#1], Aisi Fu[#2], Ben Hu[#2], Yongqing Tong[#1], Ran Liu[#2], Jiashuang Gu[3], Jianghao Liu[3], Wen Jiang[3], Gaigai Shen[2], Wanxu Zhao[2], Dong Men[4], Zixin Deng[1,5,6], Lilei Yu[7], Yan Li*[1], Tiangang Liu*[2,6]

[1] Department of Clinical Laboratory, Renmin Hospital of Wuhan University, Wuhan, 430060, China

[2] Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education and Wuhan University School of Pharmaceutical Sciences, Wuhan, 430071, China

[3] Wuhan Dgensee Clinical Laboratory Co., Ltd., Wuhan, 430075, China

[4] Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, Hubei Province, P.R. China

[5]State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200030, China

[6]Hubei Engineering Laboratory for Synthetic Microbiology, Wuhan Institute of Biotechnology, Wuhan, 430075, China

[7]Department of Internal Medicine, Renmin Hospital of Wuhan University. Wuhan, 430060, China

# These authors contributed equally to this work.

Corresponding author:

Tiangang Liu, (liutg@whu.edu.cn)； Yan Li, (yanlitf1120@163.com)

Abstract

The ongoing novel coronavirus pneumonia COVID-19 outbreak in Wuhan, China, has engendered numerous cases of infection and death. COVID-19 diagnosis relies upon nucleic acid detection; however, current recommended methods exhibit high false-negative rates, low sensitivity, and cannot identify other respiratory virus infections, thereby resulting patient misdiagnosis and impeding epidemic containment. Combining the advantages of target amplification and long-read, real-time nanopore sequencing, we developed nanopore target sequencing (NTS) to detect SARS-CoV-2 and other respiratory viruses simultaneously within 6–10 h. Parallel testing with approved qPCR kits of SARS-CoV-2 and NTS using 61 nucleic acid samples from suspected COVID-19 cases confirmed that NTS identified more infected patients as positive, and could also monitor for mutated nucleic acid sequence or other respiratory virus infection in the test sample. NTS is thus suitable for contemporary COVID-19 diagnosis; moreover, this platform can be further extended for diagnosing other viruses or pathogens.

Introduction

An ongoing novel coronavirus pneumonia (COVID-19) outbreak originating in Wuhan, China in December 2019 has subsequently spread across China and worldwide, resulting in numerous cases of infection and death[1]. Usually, COVID-19 has an incubation period of 2–7 days[2] with no obvious symptoms, during which time the virus can spread from infected to uninfected individuals. Therefore, early accurate diagnosis and isolation of patients is key to controlling the epidemic.

Nucleic acid detection is the golden standard for COVID-19 diagnosis. Real-time reverse transcription-polymerase chain reaction (qPCR) is the most recommend testing method for detecting the causative virus, SARS-CoV-2[3]. qPCR is specific, rapid, and economic, but cannot precisely analyze amplified gene fragment nucleic acid sequences; thus, positive infection is confirmed by monitoring one or two sites (depending on manufacturer guidelines). However, qPCR exhibits high false-negative rates and low sensitivity in clinical application[4], with only 30–50% positive detection ratio. False-negatives facilitate epidemic spread through delayed patient isolation and treatment, and patients mistakenly considered uninfected or cured following misdiagnosed treatment results. Another recommend detection method, sequencing, is widely applied for pathogen identification and monitoring virus evolution[5, 6] including SARS-CoV-2[7], but requires expensive equipment, operator expertise, and > 24 h turnaround time, rendering it unsuitable for the current crisis.

Several intelligent methods for RNA virus detection have developed including combining toehold switch sensors[8], which can bind to and sense virtually any RNA sequence, with paper-based cell-free protein synthesis for Ebola and Zika virus detection[9, 10], and the SHERLOCK method based on CRISPR/Cas13a for Zika or Dengue virus detection[11]. A rapid SHERLOCK method with visual results can detect SARS-CoV-2[12] and toehold switch biosensors could theoretically be

developed for rapid and high-throughput SARS-CoV-2 detection. However, the requirement of specific RNA regions as targets may negatively affect detection rates because target region mutation may limit target availability.

Moreover, pneumonia and fever may be caused by other respiratory viruses[13]. Cross-infection during the diagnosis process both spreads SARS-CoV-2 and subjects COVID-19 patients to other respiratory viruses. In severe cases, comprehensive analysis of infecting viruses is necessary. Therefore, a rapid, accurate, and comprehensive detection method is needed to inform clinical treatment and control cross-infection to reduce mortality.

Currently, COVID-19 infection and death rates in Hubei province are the highest in China. Being located at the center of this epidemic, we developed a nanopore target sequencing (NTS) method combining the advantages of target amplification and long-read, real-time nanopore sequencing for detecting SARS-CoV-2 with higher sensitivity than standard qPCR, simultaneously with other common respiratory viruses and mutated nucleic acid sequence within 6–10 h.

Results

NTS design for SARS-CoV-2 detection. NTS is based upon amplification of 11 virulence-related and specific gene fragments (*orf1ab*) of SARS-CoV-2 using a primer panel developed in-house, followed by sequencing the amplified fragments on a nanopore platform. To enhance sensitivity, we focused on virulence-related genes as targets without limitation to the sites currently recommended by Chinese or American Centers for Disease Control (CDC) in qPCR methods (Fig. 1). Because this method can precisely determine nucleic acid sequences, positive infection can be confirmed by analyzing output sequence identity, coverage, and read number.

To realize detection of pivotal SARS-CoV-2 virulence genes, we focused on the virulence region (genome bp 21,563–29,674; NC_045512.2), encoding S (1273 amino acids; AA), ORF3a (275 AA), E (75 AA), M (222 AA), ORF6 (61 AA), ORF7a (121 AA), ORF8 (121 AA), N (419 AA), and ORF10 (38 AA) proteins. We also considered the RNA-dependent RNA polymerase (RdRP) region in *orf1ab* (Fig. 1). For the virulence regions, 11 fragments of 600–950 bp were designed as targets, fully covering the 9,115 bp region (Fig.1), amplified by 22 specific primers designed considering primer-primer interaction and annealing temperature, and potential non-specific binding to human and common bacterium and fungi genomes. To improve the sensitivity *orf1ab* region amplification, we designed two pairs of nested primers to amplify 300–500 bp regions to avoid amplification failures owing to site mutation. Finally, the 26 primers were combined to develop the SARS-CoV-2 primer panel (Supplementary Table 1).

For sequencing, we chose a nanopore platform that could sequence long nucleic acid fragments and simultaneously analyze the data-output in real-time. This allowed confirmation of SARS-CoV-2 infection within a few minutes after sequencing by mapping the sequence reads to the SARS-CoV-2 genome and analysis of output sequence identity, coverage, and read number. Moreover, the accurate nucleic acid sequence generated using our pipeline could indicate whether the virulence-related genes were mutated during virus spreading, thereby rapidly providing information for subsequent epidemiological analysis. Additionally, as the MinION nanopore sequencer is portable, NTS is also suitable for front-line clinics.

NTS results interpretation and limit of detection (LoD). To test the SARS-CoV-2 detection efficiency by NTS, we used standard plasmids harboring COVID-19 virus *S* and *N* genes to simulate SARS-CoV-2. Standard plasmids were individually spiked into background cDNA

106 samples (cDNA reverse-transcribed from an uninfected respiratory flora throat swab) at 10, 100,
107 500, 1000, and 3000 copies/mL. NTS for all test samples was performed on one MinION sequencer
108 chip. Sequence data were evaluated at regular intervals using our in-house bioinformatics pipeline.
109 By mapping output reads on the SARS-CoV-2 genome, all reads with high identity were calculated
110 for each plasmid concentration. For 10 min and 1 h sequencing data, reads mapped to SARS-CoV-2
111 significantly differed from those of negative controls in all replicates at concentrations ranging from
112 3000 to 500 (Fig. 2a), and 3000 and 10 (Fig. 2b) copies/mL, respectively. This result confirmed that
113 high-copy samples could rapidly yield sufficient valid sequencing data for diagnosis, and by
114 extending the sequencing time, valid sequencing data could also be obtained from low-copy
115 samples. Notably, as more sequencing data could be achieved with additional sequencing time
116 (Supplementary Fig. 1) and clinical samples may exhibit higher complexity, thus, 10 min (for quick
117 detection) and 4 h (for final evaluation) sequencing times were used in subsequent evaluation of
118 NTS in clinical samples.

119 Evaluation of the target distribution of these valid data revealed that in higher copies samples
120 (1000 and 3000 copies/mL), all targeted regions could be detected (Fig. 2c, d). However, in lower
121 viral concentration samples (from 10 to 500 copies/mL), some targeted regions were lost (i.e., no
122 reads mapped; Fig. 2c, d), indicating that for low-quality or low-abundance samples, comprehensive
123 fragment amplification is difficult. Therefore, for accurate results, NTS cannot label a sample as
124 positive for infection by monitoring only one or two sites, as is customary for qPCR; rather, the
125 results from all target regions should be considered.

126 Accordingly, we determined a scoring rule by referring to previous judgment rules[14-16]. Firstly,
127 we counted the number of output reads with high identity to the SARS-CoV-2 genome, indicative
128 of high credibility for identification as SARS-CoV-2. By calculating the ratio of the counted valid

read numbers of the test sample to those of the negative control (with "0" in the negative control calculated as "1"), we defined that a ratio of ≥10 indicates a positive result for that fragment, scoring 1; ≥3 to 10 fold is inconclusive, scoring 0.4; and <3 is negative, scoring 0. Scores were summed to obtain the NTS score. We considered that a sample in which at least 50% fragments (6 fragments) are inconclusive or 2 fragments are positive (comparable to qPCR results) could be defined as a positive infected sample (e.g., NTS score >2.4); 3–6 inconclusive fragments or 1 positive fragment indicated a highly suspect (inconclusive) sample (e.g., NTS score of 1.2–2.4); and < 3 inconclusive or no positive fragments could be defined as negative sample (NTS score <1.2).

To determine the NTS LoD, we used the defined rules to evaluate each replicate in the simulated test. As the standard plasmids contain only 6 designed fragments (half of 12 designed fragments for SARS-CoV-2), we defined the scoring as NTS score >1.2 indicates positive detection, 0.6–1.2 is inconclusive, and < 0.6 reflects negative detection. We calculated the score of the lowest concentration (10 copies) at different times according to this scoring method and judged the positive detection rate. The results (Supplementary table 2) showed that 3/4 of the 10 copies of the standard plasmids can be judged positive from 1 h. This result is consistent with the significant comparation (Fig. 2b), that the data for 10 copies standard plasmids is also significantly different from the negative control from 1h. This result shows that our scoring system is reliable for evaluating NTS test results, and the LoD (3/4 replicates positive) was determined as 10 copies/mL with 1h sequencing data (1,372 to 43,967 reads per sample in a run with 24 samples).

SARS-CoV-2 detection using qPCR vs NTS. We performed clinical sample testing at the first-line hospital in Wuhan as soon as NTS method was established (Fig. 3). To verify NTS sensitivity, we evaluated 45 nasopharyngeal swab samples from outpatients with suspected COVID-19 early in

the epidemic. On February 6 and 7, 2020, we parallel tested these 45 samples in two batches using NTS (two chips) and qPCR (kit 2; Fig. 1). The 4 h sequencing output data (Fig. 4a) revealed that all 19 samples defined as positive by qPCR were recognized SARS-CoV-2-infected by NTS, indicating good inter-test concordance. Among 15 qPCR-inconclusive samples, 11 were recognized as SARS-CoV-2-infected, 3 as negative, and 1 inconclusive by NTS. Among 11 qPCR-negative samples, 4 were recognized SARS-CoV-2-infected, 4 as inconclusive, and 3 as negative by NTS. Overall, NTS identified a total of 34 positive samples in 45 suspected samples, 15 more than qPCR. Evaluation of output data after 10 min sequencing (Supplementary Fig. 2) revealed that 21 of 45 suspected samples were recognized as SARS-CoV-2-infected by NTS. For these samples, the 10 min and 4 h results were comparable, indicating that NTS could rapidly detect many positive samples.

However, as the tested 45 samples were from early outpatients without detailed records, suitable clinical data, such as chest computed tomographic scans, were not available to support the results. Therefore, we next evaluated samples retained from hospitalized patients with confirmed COVID-19 subjected to qPCR testing (kit 1, Fig. 1) on February 11 and 12, 2020. We randomly selected 16 samples for NTS testing on February 20, 2020. 4 h sequencing (Fig. 4b) identified all 16 positive samples, whereas only 9 samples were positive by qPCR. At the time of this writing, among these 7 samples that qPCR negative or inconclusive, electronic records indicated that subsequent qPCR testing of 4 of these 7 patients revealed two (R04 and R09) as positive whereas two (R06 and R07) remained inconclusive. This result confirmed that NTS could identify more positive cases than qPCR. Three positive samples were identified by 10 min sequencing (Supplementary Fig. 2), indicating that NTS could rapidly detect positive samples with high concentration of virus.

Evaluation of the positive target distribution for each sample (Fig. 4) showed that samples positive by both NTS and qPCR appeared to have higher nucleic acid quality or abundance, because NTS yielded more positive fragments. For qPCR-inconclusive samples, NTS yielded few, scattered positive target fragments, suggesting that low sample nucleic acid quality or abundance rendered it difficult to draw clear conclusions by qPCR based on evaluation of only two sites. Moreover, contamination of individual viral fragments did not affect NTS results. For example, although the negative control of the first chip in Fig. 4 appears to have been contaminated with a fragment containing the *N* gene, we could exclude the contamination using a high threshold, and/or base the final conclusions on the 11 remaining sites. However, negative control contamination in qPCR would invalidate the results of the whole experimental batch.

SARS-CoV-2 mutation analysis. Mutation screening of 19 samples from outpatients indicated as infection-positive by both NTS and qPCR identified single nucleotide mutations at seven sites in four samples (Table 1), three of which (S_519 of C1, N_822 of C2, and S_2472 of E3) harbored silent mutations. One variant (ORF8_251: T→C), encoding an AA change from Leu to Ser, was identified in the three samples. The ORF8_184 mutation in sample E3 also reflected a Val to Leu missense mutation. Comparison with the 67 complete SARS-CoV-2 genomes reported in the GISAID database prior to February 8, 2020 revealed that ORF8_251 contained C in 20, T in 48, and Y in 1 genome, indicating its diversity in different strains. Additionally, single genomes contained C or S at ORF8_184 whereas the remaining 67 had G, indicating that despite some inter-strain diversity, G→C transversion was rare. The remaining silent variants were not identified in the GISAID database, suggesting that the virus may harbor mutations as yet uncharacterized by existing genome-wide sequencing methods.

NTS panel for respiratory virus identification. The inability of current clinically utilized SARS-CoV-2 qPCR kits to identify the species of co-infecting viruses combined with the high false-negative rate of qPCR compromises early patient triage, resulting in wasted urgent medical resources and enhancing potential cross-contamination during the diagnosis process. Distinguishing different types of respiratory viral infections has attracted worldwide attention.

To extend the scope of NTS-based virus detection, we designed a respiratory virus primer panel for amplification of 10 respiratory viruses including bocavirus, rhinovirus, human metapneumovirus, respiratory syncytial virus, coronavirus, adenovirus, parainfluenza virus, influenza A virus, influenza B virus, and influenza C virus. We collected target gene candidates utilized for virus identification in the literature, then collected all complete and partial target gene sequences for these viruses available at GenBank (through November 1, 2019). Though multiple nucleic acid sequence alignment of each gene, the conserved regions were chosen as candidate regions for amplification. Using similar constraints as for SARS-CoV-2 target region selection, we chose 20 target amplification regions (300–800 bp) for these 10 respiratory viruses (Supplementary Table 3) capable of accurately distinguishing virus in addition to identifying virus species. We designed 59 primers including some nested primers for amplification of these regions, comprising the respiratory virus primer panel (Supplementary Table 4).

To verify the performance of this panel in NTS, we selected five virus-positive samples (influenza A virus, influenza B virus, parainfluenza, respiratory syncytial virus, and rhinovirus), the viruses in the clinical throat swabs were previously confirmed using a China Food and Drug Administration (cFDA) approved kit (Health Gene Technologies, China) based on multiplex PCR and capillary electrophoresis analysis. The five samples were mixed to create a mock virus

community and used to test the NTS virus detection capacity. NTS 10 min sequencing data (Supplementary Table 5) successfully detected four of five viruses (influenza A virus, influenza B virus, respiratory syncytial virus, and rhinovirus); the remaining one virus with lower load could be detected with 2 h sequencing data, confirming the suitability of NTS with the respiratory virus primer panel for identification of other respiratory viruses.

To verify the ability of NTS to detect SARS-CoV-2 and 10 kinds of respiratory viruses in a single assay, 13 of the 45 suspected COVID-19 outpatient samples were subjected to simultaneous detection analysis. Five replications of the plasmid containing the SARS-CoV-2 *S* and *N* genes served as the positive control and Tris-EDTA (TE) buffer was used as the negative control (in duplicate). For each sample, cDNA samples were separately amplified using the respiratory virus and the SARS-CoV-2 primer panels, then all amplified fragments were pooled. After the addition of barcodes, amplified fragments from all 20 samples (13 cases, 7 controls) were subjected nanopore sequencing on one chip. Analysis of the results (Table 2) revealed that E11 was co-infected by influenza A virus H3N2 and SARS-CoV-2.

Discussion

Herein, we developed an NTS method able to simultaneously detect SARS-CoV-2 and 10 additional types of respiratory viruses within 6–10 h, at LoD of 10 copies/mL with at least 1 h sequencing data. The detection region of SARS-CoV-2 was composed of 12 fragments covering nearly 10 kb of the genome, resulting in markedly higher sensitivity and accuracy than those of qPCR kits currently in clinical use. Notably, 22 of 61 suspected COVID-19 samples that were negative or inconclusive by qPCR testing were identified as positive by NTS. Moreover, NTS

243 enabled the detection of virus mutations; in particular, we detected a nucleotide mutation in SARS-
244 CoV-2 that was undetected in the genomic data in the current GISAID database. Although this was
245 a silent mutation, its presence suggests that the virus may have undergone mutation during the
246 spreading process. Additionally, NTS was verified as capable of detecting all five pre-added
247 respiratory viruses in a single detection. This method also detected a co-infected case (SARS-CoV-
248 2 and human influenza A virus H3N2) using a clinical specimen, illustrating the ability of NTS to
249 detect and distinguish respiratory viruses. Together, our findings indicate that NTS is highly
250 suitable for the detection and variation monitoring of current COVID-19 epidemics, directly from
251 clinical samples with same-day turnaround of results.

252 At the time of this writing, the COVID-19 epidemic remains very severe. Accurate, rapid, and
253 comprehensive nucleic acid detection methods are needed to allow patients with suspected infection
254 to be isolated and treated as soon as possible, and to accurately confirm whether the patient is cured,
255 to prevent continued epidemic spread caused by misdiagnosis. The LoD of NTS was shown to be as
256 low as 10 copies/mL, rendering it 100-fold more sensitive than qPCR (e.g., some kits describe
257 LoDs of 1000 copies/mL) and thus likely to decrease the high false-negative rate plaguing current
258 detection methods. In addition, the detection of co-infection may allow the prevention of disease
259 progression from mild to severe or might be useful to inform clinical treatment. Overall, NTS
260 combines sensitivity, broad detection range, same-day rapid turnaround time, variation monitoring,
261 and low cost (compared with whole-genome sequencing), making it the most suitable method for
262 the detection of suspected viral infections that cannot be effectively diagnosed by other methods.
263 Moreover, the MinION, the smallest Oxford Nanopore sequencer, is smaller than a cellphone; when
264 equipped with a laptop computer for data processing, it thus allows rapid performance of NTS in

various environments with low equipment cost. For data analysis, cloud analysis may also be introduced for high-throughput detection[17, 18].

Several limitations of the current NTS method should be noted. Because the designed amplified fragments are 300–950 bp in length, which constitute suitable lengths for detection by a nanopore platform as nucleic acid fragments < 200 bp cannot be readily detected[19, 20], thereby, the sensitivity of NTS for detecting target COVID-19 fragments in highly degraded nucleic acids may be hampered. Additionally, although the turnaround time of NTS is longer than that of qPCR or other possible nucleic acid detection methods (e.g., SHERLOCK[12]), 6–10 h is considered acceptable for clinical use; moreover, NTS is already the fastest strategy based on sequencing methods to date and can detect variations directly from clinical samples. Whereas the detection throughput of NTS is not high at present, the NTS method can be integrated into widely used automated or semi-automated platforms to improve the detection throughput in the future[21-23]. In addition, because PCR is included in NTS, processes involving opening the lid of the PCR tubes may cause mutual contamination between samples[24, 25]. However, this situation also is inevitable in current nucleic acid detection methods (e.g., qPCR) or other nucleic acid detection schemes (e.g., SHERLOCK[11, 12] or toehold switch biosensor[9, 10]) that also involve PCR. The introduction of integration systems or sealed devices such as microfluidics may avoid this situation[26, 27]. At present, our processes of sequencing data analysis and interpretation of results are not mature; nevertheless, as the number of test samples increases, additional test results will be collected and the process continuously optimized to obtain more accurate results.

Notably, the comparison of NTS and qPCR results indicated a high false-negative rate in the latter. This result highlights the need for extreme vigilance, as patient misdiagnosis (including patients admitted and discharged) will lead to spread of the epidemic and greater public health

threat. Suspected or negative results reported by the current qPCR methods should be subjected to a more accurate method for secondary confirmation; for this, we consider NTS as the most recommended solution currently available. The situation of co-infection, which has been reported in our previous study[13], also warrants continued attention. Based on the current centralized treatment strategy, the lack of screening for multiple viruses may lead to large-scale cross-contamination and confound clinical diagnosis and treatment. Alternatively, NTS represents and effective strategy that can rapid and accurate distinguish SARS-CoV-2 and multiple respiratory viruses at both the species and subtype level, and could be applied as a spot check in centralized clinics. Finally, the NTS method for respiratory virus detection might be extended to detect more viruses and other pathogens through the design of additional primer panels.

Methods

Primer panel design for SARS-CoV-2. The SARS-CoV-2 primer panel was designed to simultaneously detect virus virulence- and infection-related genes and variants thereof. The 21,563–29,674 bp genome region, containing the genes encoding S, ORF3a, E, M, ORF6, ORF7a, ORF8, N, and ORF10, was selected as a template to design a series of end-to-end primers. The region encoding ORF1ab was selected as a template to design a nested primer for higher sensitivity detection of SARS-CoV-2. All primers were designed using online primer-blast (https://www.ncbi.nlm.nih.c/tools/primer-blast/) and the specificity of all primers was verified against *Homo sapiens*, fungi, and bacteria. Finally, we downloaded and selected *N*, *S*, *rdrp*, and *E* gene sequences of SARS-related viruses available at GenBank through January 1st, 2020 (accession NC_045512). Multiple sequence alignment of SARS-CoV-2 against SARS-related viruses was performed using Clustal W (version 1.83) for each gene individually and the alignment was used for

311 in-silico evaluation of specificity of the designed primers to SARS-CoV-2. All the specific primers
312 were collected to form the SARS-CoV-2 primer panel.
313
314 **Primers panel design for 10 kinds of respiratory virus detection**. The target genes for each virus
315 were selected based on previous literature retrieval and all complete and partial gene sequences
316 available in GenBank through November 1, 2019 were downloaded. The list for each target gene
317 was manually checked and artificial sequences (e.g., lab-derived, synthetic) along with sequence
318 duplicates was removed, resulting in a final list. Multiple sequence alignment was performed using
319 Clustal W (version 1.83) for each gene individually and the variation rate of each base was
320 calculated using an in-house pipeline. The final primers for each virus were manually selected
321 following the previous metrics[28] for multiplex PCR design with an expected amplicon length range
322 from 300 to 800 bp.
323
324 **LoD of the NTS test**. Individual NTS libraries were prepared from a virus-negative nasopharyngeal
325 swab spiked with plasmids containing synthetic *S* and *N* genes of COVID-19 at concentrations of 0,
326 10, 100, 500, 1,000, and 3,000 copies/mL, with four replicates at each concentration. The NTS
327 libraries were prepared as described above and sequenced using MinION for 10 min, 30 min, 1 h, 2
328 h, and 4 h. The sequencing data were processed as described for virus identification. The LoD was
329 determined when the concentration of reads mapped to COVID-19 was significantly higher than
330 that for the negative control in 3/4 replicates.
331
332 **NTS detection method**. The targeted genes were amplified using the SARS-CoV-2 or 10
333 respiratory virus primer panel in a 20 μL reaction system with 5 μL total nucleic acid, 5 μL primer

(10 μM), and 10 μL 2× Phusion U Multiplex PCR Master Mix (Thermo Fisher, USA)[29, 30]. Amplification was performed in a C1000 Thermocycler (Bio-Rad, USA) using the following procedure: 1 cycle at 94 °C for 3 min and 30 cycles at 95 °C for 10 s, 55 °C for 50 s, and 68 °C for 5s, followed by a final elongation step at 68 °C for 5 min. The product of the first-step was purified with 0.8× AMpure beads (Beckman Coulter, USA) and eluted in 10 μL Tris- EDTA (TE) buffer. Then, 5 μL eluate was used for second-step PCR with 5 μL barcoded primer (10 μM) and 10 μL 2× Phusion U Multiplex PCR Master Mix. The barcode sequence was from the Nanopore PCR barcode kit (EXP-PBC096; UK) and all primer oligos and full-length *S* and *N* gene fragments were synthesized by Genscript (China). The products of second-step PCR from the different samples were pooled with equal masses. TE buffer was assayed in each batch as a negative control. Sequencing libraries were constructed using the 1D Ligation Kit (SQK-LSK109; Oxford Nanopore, UK) and sequenced using Oxford Nanopore MinION or GridION.

Nanopore sequencing data processing. Basecalling and quality assessment for MinION sequencing data were performed using high accuracy mode in Guppy (v. 3.1.5) software; for GridION, the process was conducted using MinKNOW (v. 3.6.5) integrated in the instrument. Sequencing reads with low quality and undesired length were discarded. Then, Porechop[31] (v. 0.2.4) was used for adaptor trimming and barcode demultiplexing for retained reads.

Mapping tool and mapping database. BLASTn[32] (v. 2.9.0+) was used to map the reads of each sample against the virus genome reference database. All virus genomic sequences were downloaded from NCBI Refseq FTP and the SARS-CoV-2 genome sequence was added to the BLAST database

subsequently. The taxonomy of each read was assigned according to the taxonomic information of the mapped subject sequence.

Sequence correction and candidate mutation calling. Sequence correction was performed using medaka[33] (v. 0.10.5), which is a tool to create a consensus sequence of nanopore sequencing data. For each target sequencing region, 30 consensus sequences were generated using medaka's default settings through the medaka_consensus program. Subsequently, the consensus sequences were aligned to the reference sequence of target sequencing regions using the multiple sequence alignment tool ClustalW[34] (version 1.83). The variants within certainty regions (except sequence homopolymeric regions and primer binding sites)[35] and with appropriate coverage (covered by at least 90% consensus sequences and at least 50% uncorrected reads) were accepted as candidate nucleotide mutations.

Interpretation of NTS results. The sequenced data were obtained at regular intervals after sequencing, then filtered to obtain valid reads. For determining whether the target was positive, interpretation was performed using the previous rule with modification[14-16]. In brief, if the read matches the design fragment, the read will be counted. The mapping score was determined as 1, 0.4, or 0 when the ratio of count number in the sample to the negative control of each target was > 10, between 3–10, or < 3. The total mapping score of each target was summed and samples with > 2.4 total mapping score were defined as positive for SARS-CoV-2 infection; 1.2 to 2.4 total mapping score indicated an inconclusive result, and < 1.2 total mapping score was considered to indicate negative for infection. For determination of the other 10 kinds of common respiratory virus, a

sample was considered positive for the virus if positive for at least one designed site, otherwise it was negative.

Total nucleic acid extraction from clinical specimens. Clinical throat swab specimens were collected in 10 mL of Viral Transport Medium (Becton Dickinson, USA) from 45 suspected COVID-19 outpatients, and 16 hospitalized patients with COVID-19 at Renming Hospital of Wuhan University in Wuhan during February 2020. All throat swabs were sent to a clinical laboratory and processed immediately. Swabs were vortexed in 1 mL of TE buffer and centrifuged at 20,000 × *g* for 10 min. The supernatant was removed and 200 μL of the specimen was retained for total nucleic acid extraction. Total nucleic acid was extracted from 200 μL of pre-treated samples using the Sansure SUPRall DNA Extraction Kit (Changsha, China) following the manufacturer's instructions. Extracted total nucleic acid was stored at 70 °C until qPCR or NTS testing.

qPCR for confirmation of SARS-Cov-2 infection. The total isolated nucleic acid was used for qPCR assay following the manufacturer's instructions. Briefly, qPCR was carried out in a 25 μL reaction system using a novel coronavirus qPCR kit (kit 1, Huirui, China) with 5 μL total nucleic acid, or 20 μL reaction system using the 2019-nCoV qPCR kit (kit 2, BioGerm, China) with 5 μL total nucleic acid. For kit 1, amplification was performed using a Quantstudio Dx Real-time PCR system (Thermo Fisher, USA) with the following procedure: 1 cycle at 50 °C for 15 min and 95 °C for 5 min, and 35 cycles at 95 °C for 10 s, 55 °C for 40 s. The FAM and ROX fluorescence channels were used to detect *Orf1ab* and *N* gene, respectively. Successful amplification of both genes and Ct value ≤35 was recognized as positive for SARS-CoV-2 infection; Ct value between

35.2 to 39.2 was recognized as inconclusive, and one of the genes being undetected or Ct ≥ 39.2 was recognized as negative. For kit 2, amplification was performed in a CFX96 Thermocycler (Bio-Rad) using the following procedure: 1 cycle at 50 °C for 10 min and 95 °C for 5 min, and 35 cycles at 95 °C for 10 s, 55 °C for 40 s. The FAM, HEX, and CY5 fluorescence channels were used to detect *Orf1ab*, *E*, and *N* genes, respectively. This kit only utilized the results of the *Orf1ab* and *N* gene to reach a conclusion. Successful amplification of both genes and Ct value ≤35 was recognized as positive for SARS-CoV-2 infection; only one site with Ct value ≤35 or both genes between Ct 35 to 38 was taken as inconclusive, and no successful amplification or Ct ≥ 38 was recognized as negative for infection.

Clinical records of patients. The clinical records of patients were kept in Renmin Hospital of Wuhan University. Clinical, laboratory, and radiological characteristics and treatment and outcome data were obtained using data collection forms from electronic medical records. The data were reviewed by a trained team of physicians. The study and use of all records were approved by the Ethics Committee of Hubei Provincial Renmin Hospital (WDRY2019-K056).

Data availability

All data for support the study result are included in this published article (and its supplementary information files). Other data generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Chinese CDC. Distribution of novel coronavirus pneumonia. http://2019ncov.chinacdc.cn/2019-nCoV/
2. Guan, W.-j. et al. Clinical characteristics of 2019 novel coronavirus infection in China. *N Engl J Med*, (2020).

doi: 10.1056/NEJMoa2002032.

3. Corman, V.M. et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro surveill* **25** (2020).
4. Li Jin, et al. Analysis of false-negative results for 2019 novel coronavirus nucleic acid test and related countermeasures. *Chin J Lab Med* **43** (2020).
5. Wilson, M.R. et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med* **370**, 2408-2417 (2014).
6. Metsky, H.C. et al. Zika virus evolution and spread in the Americas. *Nature* **546**, 411-415 (2017).
7. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* (2020).
8. Green, A.A., et al. Toehold switches: de-novo-designed regulators of gene expression. *Cell* **159**, 925-939 (2014).
9. Pardee, K. et al. Paper-based synthetic gene networks. *Cell* **159**, 940-954 (2014).
10. Pardee, K. et al. Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* **165**, 1255-1266 (2016).
11. Gootenberg, J.S. et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438-442 (2017).
12. Feng Z. et al. A protocol for detection of COVID-19 using CRISPR diagnostics. https://www.broadinstitute.org/files/publications/special/COVID-19%20detection%20(updated).pdf
13. Wang, M. et al. Clinical diagnosis of 8274 samples with 2019-novel coronavirus in Wuhan. Preprint at https://www.medrxiv.org/content/10.1101/2020.02.06.20020974v1
14. Blauwkamp, T.A. et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol* **4**, 663-674 (2019).
15. Langelier, C. et al. Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic Cellular Transplant Patients. *Am J Respir Crit Care Med* **197**, 524-528 (2018).
16. Metsky, H.C. et al. Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. *Nat Biotechnol* **37**, 160-168 (2019).
17. Charalampous, T. et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat Biotechnol* **37**, 783-792 (2019).
18. Greninger, A.L. et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med* **7**, 99 (2015).
19. Wei, S., Weiss, Z.R. & Williams, Z. Rapid Multiplex Small DNA Sequencing on the MinION Nanopore Sequencing Platform. *G3 (Bethesda)* **8**, 1649-1657 (2018).
20. Wilson, B.D., Eisenstein, M. & Soh, H.T. High-Fidelity Nanopore Sequencing of Ultra-Short DNA Targets. *Anal Chem* **91**, 6783-6789 (2019).
21. Kong, N. et al. Automation of PacBio SMRTbell NGS library preparation for bacterial genome sequencing. *Stand Genomic Sci* **12**, 27 (2017).
22. Mardis, E. & McCombie, W.R. Automated Library Preparation for DNA Sequencing. *Cold Spring Harb Protoc* **2017** (2017).
23. Quail, M.A., Gu, Y., Swerdlow, H. & Mayho, M. Evaluation and optimisation of preparative semi-automated electrophoresis systems for Illumina library preparation. *Electrophoresis* **33**, 3521-3528 (2012).
24. Eisenhofer, R. et al. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol* **27**, 105-117 (2019).
25. Martin, P., Laupland, K.B., Frost, E.H. & Valiquette, L. Laboratory diagnosis of Ebola virus disease. *Intensive Care Med* **41**, 895-898 (2015).
26. Yager, P. et al. Microfluidic diagnostic technologies for global public health. *Nature* **442**, 412-418 (2006).
27. Wang, S., Inci, F., De Libero, G., Singhal, A. & Demirci, U. Point-of-care assays for tuberculosis: role of nanotechnology/microfluidics. *Biotechnol Adv* **31**, 438-449 (2013).

28. Shen, Z. et al. MPprimer: a program for reliable multiplex PCR primer design. *BMC Bioinformatics* **11**, 143 (2010).
29. Calus, S.T., Ijaz, U.Z. & Pinto, A.J. NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. *Gigascience* **7** (2018).
30. Gomez, C.A., Budvytiene, I., Zemek, A.J. & Banaei, N. Performance of Targeted Fungal Sequencing for Culture-Independent Diagnosis of Invasive Fungal Disease. *Clin Infect Dis* **65**, 2035-2041 (2017).
31. R., W., Edn. 0.2.4 Porechop: a tool for finding and removing adapters from Oxford Nanopore reads. (2017).
32. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).
33. Research, O., Edn. 0.10.5 medaka: Sequence correction provided by ONT Research. (ONT Research, 2018).
34. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* **22**, 4673-4680 (1994).
35. Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228-232 (2016).
36. Specific primers and probes for detection 2019 novel coronavirus. http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html
37. Information for Laboratories 2019-nCoV Requests for Diagnostic Panels and Virus. https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html

Competing Interests

Wuhan Dgensee Clinical Laboratory Co., Ltd have applied patent on this new strategy.

494 **Figure Legends**

495 **Fig. 1 | Amplification targets of the NTS and qPCR method.** NTS detected 12 fragments
496 including ORF1ab and virulence factor-encoding regions. For qPCR, the Chinese CDC
497 recommends *Orf1ab* and *N* sites as targets,[36] the United States CDC recommends three target sites
498 in the *N* gene,[37] and Corman *et al* (2020) recommend RNA-dependent RNA polymerase (RdRP) in
499 *orf1ab* and *E* sites as the targets. Kit 1 is a cFDA-approved kit with two target sites used in this
500 study; kit 2 is a cFDA-approved kit with three target sites used in this study.

501

502 **Fig. 2 | Performance verification test of NTS for detecting SARS-CoV-2 using standard**
503 **synthetic *S* and *N* genes.** Comparison of all SARS-CoV-2 reads detected by NTS in replicates with
504 different concentrations and negative controls using 10 min (**a**) or 1 h (**b**) sequencing data. Read
505 counts mapped to each target region of the SARS-CoV-2 genome in replicates with different
506 concentrations with 10 min (**c**) to 1 h (**d**) sequencing data. Two-tailed Student t-test (for normal
507 distribution samples) or Mann–Whitney U-test (for non-normal distribution samples): ns, not
508 significant, $*P < 0.05$; bars represent the means ± SD.

509

510 **Fig. 3 | NTS testing in a front-line hospital in Wuhan**

511

512 **Fig. 4 | Comparison of 61 nucleic acid sequences from clinical samples obtained using NTS (4**
513 **h) and qPCR. a**, Comparison of 45 nucleic acid sequences from samples of patients with suspected
514 COVID-19 obtained using NTS and qPCR (kit 2). **b**, Comparison of 16 nucleic acid sequences from
515 patients with confirmed disease obtained using NTS and qPCR (kit1). The numbers in the table on
516 the left represent the number of mapped reads according to our rules. PC: positive control. The

517 plasmid harboring an *S* gene was used as a positive control in NTS testing; a positive sample in the

518 kit served as a positive control in qPCR testing. NC: negative control. TE buffer was used as a

519 negative control in NTS testing; H_2O in the kit served as a positive control in qPCR testing. Pos:

520 positive. Inc: inconclusive. Neg: Negative.

521

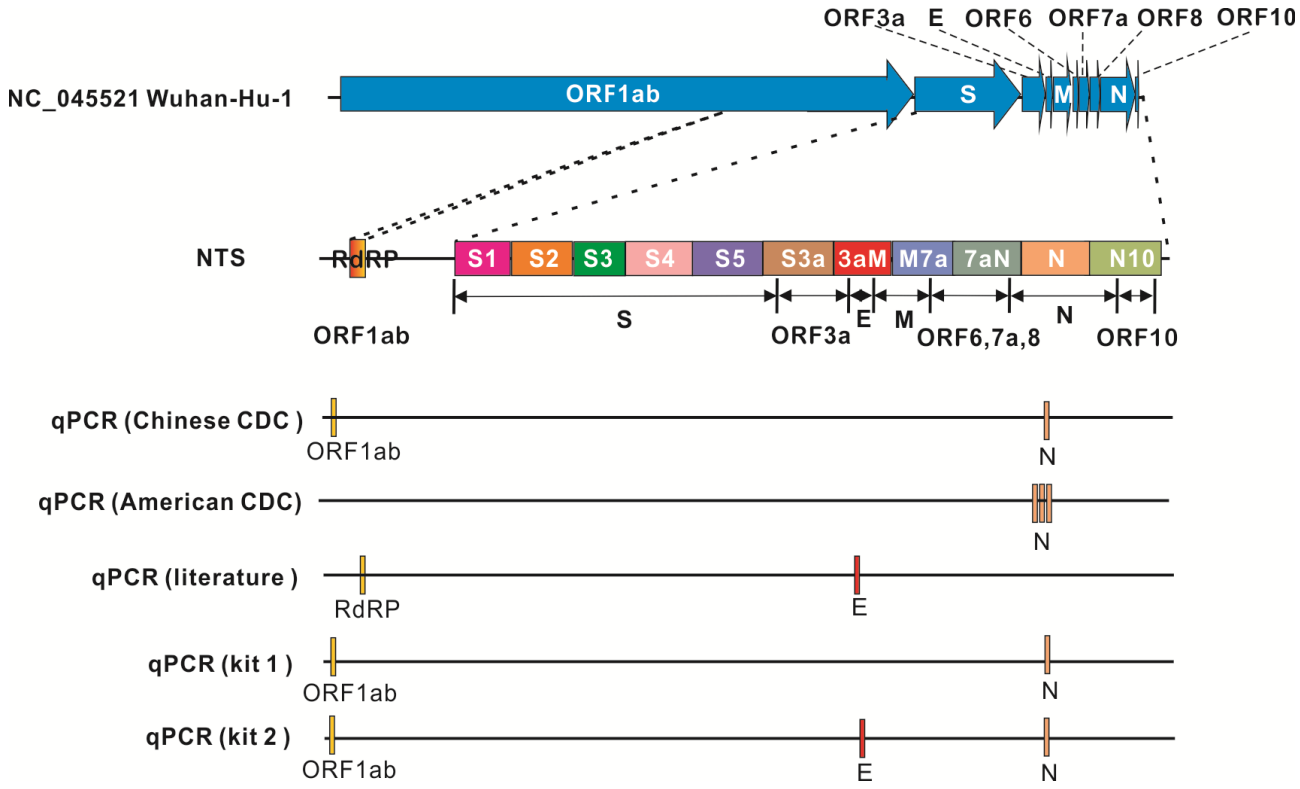

Figure 1

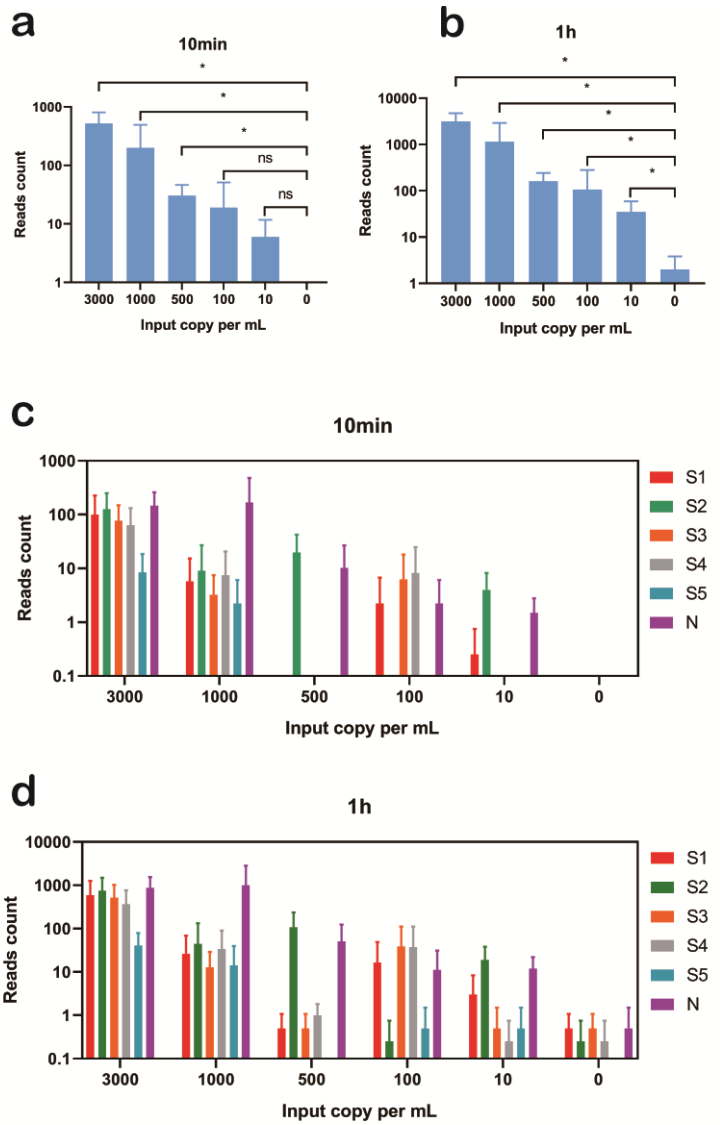

Figure 2

528

529 **Figure 3**

a

Paitent ID	NTS target sites											Final result of NTS and qPCR		
	RdRP	S1	S2	S3	S4	S5	S3a	3aM	M7a	7aN	N	N10	NTS Score	NTS	qPCR
A2	96	60	118	75	40	4	12	14	60	53	132	4	9.8	Pos	Pos
A4	797	793	1079	901	744	64	123	190	439	1179	2014	32	11.4	Pos	Pos
C1	2768	2213	2714	1882	2212	202	620	498	2538	2628	3167	364	12	Pos	Pos
C2	1680	1883	3491	1810	2445	92	737	268	1151	2973	3052	464	11.4	Pos	Pos
D12	3731	2683	4080	2392	3868	464	1709	815	2193	4825	2738	116	11.4	Pos	Pos
E5	124	211	324	121	175	9	70	24	163	359	385	3	9.8	Pos	Pos
G1	2026	791	1636	1375	1158	109	382	252	1335	2532	2921	82	11.4	Pos	Pos
H9	2015	2260	3624	1817	2691	223	647	481	1990	2621	3273	18	12	Pos	Pos
E1	582	237	258	215	180	36	54	19	237	377	360	0	10	Pos	Pos
A6	7	7	9	3	4	3	7	0	1	4	8	1	3.2	Pos	Inc
A9	11	1	40	2	1	7	0	5	20	63	55	0	4.8	Pos	Inc
A12	6	4	5	6	10	1	4	2	4	12	161	0	3.6	Pos	Inc
B4	3	6	11	6	10	1	4	1	2	8	6	1	3.4	Pos	Inc
C11	7	10	5	3	11	1	3	0	1	0	35	1	3.2	Pos	Inc
D2	0	2	8	2	1	0	2	2	4	4	6	1	0.8	Neg	Inc
F5	208	233	3	149	6	0	2	0	2	6	9	1	3.8	Pos	Inc
F12	3	1	5	1	1	1	1	0	0	1	7	0	0.4	Neg	Inc
H12	14	14	16	10	8	2	1	2	10	7	43	2	4.6	Pos	Inc
D10	3	2	5	1	1	0	0	0	1	2	2	0	0.4	Neg	Neg
A1	14	7	14	8	8	4	9	1	6	7	10	0	4.2	Pos	Neg
A3	5	10	13	2	4	1	0	1	3	2	8	0	2.2	Inc	Neg
A5	0	0	2	1	1	0	1	0	0	0	9	2	0	Neg	Neg
A7	1	8	3	1	0	1	1	0	0	22	549	41	2.4	Inc	Neg
A8	2	6	7	2	5	1	0	1	0	0	6	1	1.2	Inc	Neg
PC	0	622	504	386	517	339	26	0	0	0	0	0	6	Pos	Pos
NC	1	0	2	0	0	0	0	0	2	0	315	1	NA	Neg	Neg
A11	150	17	9	67	12	5	6	0	1	810	294	0	6.6	Pos	Pos
B9	1410	179	711	808	198	9	637	75	146	1117	1285	85	11.4	Pos	Pos
C9	1145	190	547	89	421	12	69	48	135	1317	63	0	11	Pos	Pos
D9	702	122	1440	708	769	28	807	45	74	2459	1931	0	11	Pos	Pos
D11	270	324	517	71	418	4	306	32	96	527	604	0	10.4	Pos	Pos
E3	3	2	94	37	132	15	3	0	18	5	133	1	7.2	Pos	Pos
E11	353	277	766	302	231	11	536	40	76	656	1078	157	12	Pos	Pos
F11	2009	511	1307	1166	390	30	1604	129	245	2092	2631	0	11	Pos	Pos
G11	2709	1944	2885	1735	1983	71	2334	418	595	3523	2964	8	11.4	Pos	Pos
G12	4200	1202	4765	1433	1532	6	4131	58	1711	6722	1125	0	10.4	Pos	Pos
C12	125	187	110	2	1	12	99	23	60	88	91	2	9	Pos	Inc
E6	118	123	224	2	16	1	2	16	3	3	2	1	5.8	Pos	Inc
E12	6	6	3	4	3	3	6	0	0	8	6	1	3.2	Pos	Inc
F8	5	157	5	7	295	2	5	0	1	332	239	0	5.2	Pos	Inc
G6	1	4	173	4	7	1	2	0	0	3	2	0	2.2	Inc	Inc
H3	1	0	4	26	1	1	0	0	0	0	0	0	1	Neg	Inc
A10	0	0	0	0	0	0	0	0	0	0	52	0	1	Neg	Neg
B1	1	5	7	3	6	33	3	0	0	4	8	0	3.4	Pos	Neg
B2	27	4	1	2	11	0	0	0	0	15	2	0	3	Pos	Neg
B3	21	76	10	8	118	4	8	0	14	212	288	1	7.6	Pos	Neg
B5	4	6	4	3	9	1	0	0	1	6	4	0	2.4	Inc	Neg
PC	0	916	848	588	848	593	17	0	0	0	1	0	6	Pos	Pos
NC	1	2	2	0	0	0	1	0	0	0	0	0	NA	Neg	Neg

b

Paitent ID	NTS target sites											Final result of NTS and qPCR		
	RdRP	S1	S2	S3	S4	S5	S3a	3aM	M7a	7aN	N	N10	NTS Score	NTS	qPCR
R01	1	4	38	0	2	0	3	0	0	11	11	0	3.8	Pos	Pos
R02	7	2	12	1	6	0	6	0	2	5	7	1	3	Pos	Neg
R03	50	0	17	1	3	0	10	1	0	13	8	0	4.8	Pos	Pos
R04	1	0	13	1	0	0	3	0	0	5	18	0	2.8	Pos	Neg
R05	135	0	127	23	12	0	176	1	109	580	460	0	8	Pos	Pos
R06	90	14	42	31	1	1	38	0	0	2	95	0	6	Pos	Inc
R07	0	0	4	7	12	2	1	0	0	28	9	0	3.2	Pos	Inc
R08	33	1	4	1	4	0	3	0	1	1	3	0	2.6	Pos	Inc
R09	6	2	8	2	6	0	16	0	0	7	3	1	3	Pos	Neg
R10	3	2	59	0	18	0	4	0	1	4	55	0	4.2	Pos	Neg
R11	6131	496	8613	1310	3526	38	5364	60	317	5481	5436	178	12	Pos	Pos
R12	610	100	882	120	55	9	340	18	2	898	345	0	9.4	Pos	Pos
R13	40	0	0	0	0	2	191	1	0	65	0	0	3	Pos	Pos
R14	7	0	15	1	7	0	6	0	1	6	11	0	3.6	Pos	Pos
R15	35	11	188	0	93	0	65	0	2	75	49	0	7	Pos	Pos
R16	8	0	11	0	6	0	8	0	0	3	6	1	3	Pos	Pos
NC	0	0	0	0	0	0	0	0	0	0	0	0	NA	Neg	Neg
NC	0	0	0	0	0	0	0	0	0	0	0	0	NA	Neg	Neg

Positive Inconclusive Negative

Figure 4

532 **Table 1 | Variations of SARS-CoV-2 detected by NTS**

Sample	Site (Gene_position)	Base change	Base change present in the genome[a]	Amino acid change
C1	S_519	G→A	0	NC
C2	ORF8_251	T→C	20	Leu→Ser
C2	N_822	C→T	0	NC
E3	S_2472	C→T	0	NC
E3	ORF8_251	T→C	20	Leu→Ser
E3	ORF8_184	G→C	1	Val→Leu
E5	ORF8_251	T→C	20	Leu→Ser

533 [a] The number indicates the count of genomes in which the base change appeared as reported in the

534 GISAID database prior to February 8, 2020. NC: no change.

535

536 **Table 2 | Results of NTS for detecting SARS-CoV-2 and common respiratory viruses**

Patients ID	NTS result (SARS-CoV-2)	NTS result (common respiratory viruses)
F11	Positive	ND
E11	Positive	human influenza A virus H3N2
A11	Positive	ND
B9	Positive	ND
C9	Positive	ND
D9	Positive	ND
D11	Positive	ND
C12	Positive	ND
E6	Positive	ND
B3	Positive	ND
E12	Negative	ND
G6	Inconclusive	ND
B1	Inconclusive	ND
Positive control	Positive	ND
Positive control	Positive	ND
Positive control	Positive	ND
Positive control	Positive	ND
Positive control	Positive	ND
Negative control	ND	ND
Negative control	ND	ND

537 ND: not detected

538