Testing for SARS-CoV-2 in cellular components by routine nested RT-PCR followed by DNA sequencing

Sin Hang Lee [1]*

1 Affiliation
Milford Molecular Diagnostics Laboratory, 2044 Bridgeport Avenue, Milford, CT 06460, U.S.A.

* Correspondence: shlee01@snet.net

Received: June 29, 8:12 a.m. (local Japan time); Accepted by the International Journal of Geriatrics and Rehabilitation: June 29, 23:33 p.m.; Date of initial revision by author in response to Dr. James Lyons-Weiler's review: July 12, 14:03 p.m.; Date of second revision by author in response to Dr. Harumi Kuno- Sakai's review : July 13, 12:25 p.m.; Date of the acceptance by the journal's editorial board of Dr. Sin's manuscript revised twice in response to reviews by Dr. James Lyons-Weiler and Dr. Harumi Kuno-Sakai as the finalized excellent manuscript suitable for publication: July 13, Uploaded online and Published: July 17

Abstract: Currently, molecular tests for SARS-CoV-2 infection are primarily based on reverse transcription-quantitative polymerase chain reaction (RT-qPCR) on cell-free fluid samples of respiratory tract specimens. These tests measure the rate of fluorescent signal accumulation as a surrogate for direct DNA sequence determination and are known to generate false-negative and false-positive results. The author has developed a routine protocol to test the cellular components of respiratory tract specimens instead of cell-free fluids only and to use conventional nested RT-PCR to amplify the target nucleic acid for high detection sensitivity. A 398-bp heminested PCR amplicon is used as the template for direct DNA sequencing to ensure no false-positive test results. Using this protocol to re-test 20 reference samples prepared by the Connecticut State Department of Public Health, the author found 2 positives among 10 samples classified as negative by RT-qPCR assays. One of these two positive samples contained a mutant with a novel single nucleotide insertion in the N gene and a wild-type parental SARS-CoV-2. Of the 10 samples classified as positive by RT-qPCR assays, only 7 (7/10) were confirmed to contain SARS-CoV-2 by heminested PCR and DNA sequencing of a 398-bp amplicon of the N gene. One of the latter 7 positive SARS-CoV-2 isolates belongs to a newly discovered mutant first isolated from a specimen collected in the State of New York on March 17, 2020, according to information retrieved from the GenBank database. Routine sequencing of a 398-bp PCR amplicon can categorize any isolate into one of 6 clades of SARS-CoV-2 strains known to circulate in the United States. The author proposes that extremely accurate routine laboratory tests for SARS-CoV-2 be implemented as businesses attempt to return to normal operation in order to avoid raising false alarms of a re-emerging outbreak. False-positive laboratory test reports can easily create unnecessary panic resulting in negative impacts on local economies.

Keywords: Nested RT-PCR; DNA sequencing; cellular components; SARS-CoV-2; false-negative RT-qPCR; false-positive RT-qPCR; single nucleotide insertion; 398-base; mutant; parental virus

1. Introduction

Accurate diagnosis and isolation of infectious patients without delay are the key steps in reducing the spread of emerging highly contagious diseases, like COVID-19. False-negative laboratory test results allow infected patients with mild clinical symptoms to spread SARS-CoV-2 among susceptible persons. False-positive test results may lead to placement of non-infected persons in the same isolation rooms with COVID-19 patients; eventually the non-infected individuals may become true-positive patients.

The lack of timely appropriate response from the Centers for Disease Control and Prevention (CDC) to the need of an accurate laboratory test for SARS-CoV-2 has been reported to be a factor contributing to the pandemic of COVID-19 in the United States in 2020 [1]. On February 2, 2020 the CDC was the only

place in the country that could perform SARS-CoV-2 tests. But on February 12, the CDC announced that the test was providing inconclusive results. By then, the United States had reported 11 COVID-19 cases.

The CDC designed its own test in March 2020. The Food and Drug Administration (FDA) picked a conservative testing strategy, allowing laboratories to use only the CDC test kits distributed under the Emergency Use Authorization (EUA) authority. When the CDC test kits failed by generating many false-positive results, neither a new strategy nor a new test was available for more than two weeks [1]. Eventually, the pandemic COVID-19 spread widely and rapidly throughout the country [2].

After an investigation, the lawyers from Department of Health and Human Services of the U. S. government concluded that the faulty CDC test kits were likely contaminated due to "time pressure'' [3].

However, independent scientists have reported that at least two sets of the N primers used in the CDC test kits were found to give false positive signals even in the absence of cDNA (no template control condition) [4], indicating possible faulty designs. Yet from February 4 to July 7, 2020, the FDA issued 104 individual EUAs for Molecular Diagnostic Tests for SARS-CoV-2 to device manufacturers [5]. These commercial devices are almost all based on RT-qPCR using the CDC primers and probes or using undisclosed primers and probes which cannot be independently verified.

One group of scientists in Australia tested a commercial RT-qPCR test kit and found its positive predictive value to be only 55.56%. The authors suggested that any positive results derived from one commercial test kit should be confirmed using another nucleic acid test or nucleotide sequencing [6].

Another attempt to mitigate false-positive results generated by commercial RT-qPCR kits was to develop more conventional PCR-based protocols, using the primers of the RT-qPCR kits or newly developed primers, and apply a multiplex PCR-based protocol that allowed the simultaneous testing of primer sets for RdRP, N, E, and S genes all in one reaction [7]. However, the authors offered no means to verify the various PCR products. In diagnostic virology, the optimization of multiplex PCRs can pose several difficulties, including poor sensitivity or specificity and/or preferential amplification of certain specific targets. The presence of more than one primer pair in the multiplex PCR increases the chance of obtaining spurious amplification products, primarily because of the formation of primer dimers. These nonspecific products may be amplified more efficiently than the desired target, consuming reaction components and producing impaired rates of annealing and extension [8].

False-positive RT-qPCR for SARS-CoV-2 in preoperative screening has caused unnecessary delay of urgent otolaryngology surgeries in a university hospital [9].

When a false-negative RT-qPCR for SARS-CoV-2 is generated, the tendency is to lay blame on faulty specimen collection and handling [10]. But the cause may be more complex.

The RT-qPCR test kits distributed by the CDC, also known as the CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel [11], depend on using commercial viral RNA purification devices for sample preparation. Nasopharyngeal or oropharyngeal samples should be obtained by using a Nylon flocked swab, if available, to enhance the collection and release of sufficient human cellular material to be tested.

However, these commercial viral RNA preparation kits, for example, the QIAamp® Viral RNA Mini Kit (Qiagen), are for purification of viral RNA from 180 μL of plasma, serum, cell-free body fluids and culture supernatants. Samples collected in viral transport media or saline need to be centrifuged or filtered

to remove the cellular components for proper viral RNA sample preparation [12]. The Roche DNA and Viral NA Small Volume Kit is for purifying nucleic acids from up to 200 μL sample volumes using the MagNA Pure 96 Instrument.

SARS-CoV-2 must grow and replicate in a living cell. The viral particles are primarily located in intra-cytosolic vacuoles of the infected cell [13]. The number of SARS-CoV-2 particles per infected cell has not been published. But according to the studies on other viruses, such as the human immunodeficiency virus (HIV) [14] and the human papillomavirus (HPV) [15], one infected cell in a patient's specimen may contribute several thousands of copy numbers of viral genome equivalents to be tested. Therefore, testing the cellular components of a nasopharyngeal swab rinse or bronchoalveolar lavage rather than the cell-free supernatant may raise the detection sensitivity of nucleic acid-based tests for molecular diagnosis of SARS-CoV-2 infection. An additional advantage of testing the host cells for virus is to assure that a positive result is an indication of having detected viral particles capable of causing infection, not merely non-infective free viral RNA residues floating in the extracellular fluid.

Nucleic acid-based diagnostics for infectious diseases are designed to determine the nucleotide sequence in a unique segment of the genome of the pathogen. Methodologies available to diagnostic laboratories for sequence determination include restriction fragment length polymorphism (RFLP), probe hybridization, and direct DNA sequencing.

The CDC RT-qPCR assay panel uses TaqMan® probes to bind 2 or 3 target DNA sequences if such sequences are present in the sample. The probes are labeled at the 5'-end with a fluorescent dye reporter molecule 6-carboxyfluorescein (FAM) and with the quencher, Black Hole Quencher 1 (BHQ-1) at the 3'-end. The principle of this test relies on the 5′–3′ exonuclease activity of *Taq* polymerase to cleave and degrade the dual-labeled probes which have annealed to the complementary target sequences during hybridization. Degradation of the probe releases the fluorophore from it and breaks its proximity to the quencher, thus relieving the quenching effect and allowing fluorescence of the fluorophore. It is assumed that the intensity of fluorescence detected in the qPCR thermal cycler is directly proportional to the fluorophore released and the amount of DNA template present in the PCR. However, even when a probe anneals to a partially matching sequence, enzymatic primer extension may take place with release of the fluorophore and the quencher thereafter as the probe degrades. Therefore, the fluorescent signals accumulated in RT-qPCR are merely a surrogate of the probe sequence, not the true image of the target nucleotide sequence of the template which may not exist in the sample being tested.

In contrast, automated DNA sequencing is considered the gold standard technology for nucleotide sequence determination because each fluorescent signal is emitted from a specific fluorescent dye coupled to a specific nucleotide molecule, reporting the actual position of the labeled nucleotide in a DNA sequence during capillary electrophoresis. The computer-generated sequence electropherogram represents a true image of the alignment of the 4 nucleic acid bases in the template.

According to the guidance of the U.S. Food and Drug Administration (FDA) on molecular diagnosis of viral infection caused by human papillomavirus (HPV), if the performance of a newly introduced molecular diagnostic device deviates from an FDA-approved device on the market, a conventional PCR detection followed by Sanger sequencing on both strands of the PCR amplicon (bi-directional sequencing) which contains a minimum of 100 contiguous bases is acceptable as valid diagnostics for HPV infection provided the sequence matches the reference or consensus sequence, *e.g.* with an Expected Value (E-Value) $<10^{-30}$ for the specific HPV DNA target based on a BLAST search of the GenBank database [16]. It

implies that nucleotide sequencing of a PCR amplicon of certain size with supportive GenBank BLAST search results is a *de facto* gold standard.

The DNA probes used in the CDC RT-qPCR test kits for SARS-CoV-2 assay are about 25 bases long. This design of RT-qPCR with multiple short probes does not meet the FDA requirement for nucleic acid-based molecular diagnostics for viral disease infections. In molecular diagnostics for infectious diseases when complex human specimens are tested, the significance of detecting 4 short sequences each of 25 bases long by hybridization does not equate to that of a DNA sequence composed of 100 contiguous bases that matches a signature sequence of the genome of the pathogen. Unconnected short DNA fragments may come from different sources in a complex human specimen.

The crucial step in most nucleic acid-based assays is to amplify the target DNA or cDNA in a complex clinical sample. PCR amplification always faces problems caused by inhibitors carried over from the clinical sample. Real-time PCR or qPCR is no exception. No commercial viral RNA extraction kits can remove all PCR inhibitors from the human respiratory tract specimens. For example, the QIAamp Viral RNA Mini Kit is not designed to separate viral RNA from cellular DNA [12]. The CDC test panel [11] also requires demonstration of human DNA in the test samples. Non-target nucleic acid molecules and other PCR inhibitors are usually co-extracted and co-purified along with SARS-CoV-2 RNA [17]. Inhibition of PCR by host DNA is a well-recognized problem in molecular diagnosis of bacterial infections [18]. These PCR-inhibitory molecules may come from the host cells, and from other viruses, bacteria and fungi normally residing in the human respiratory tract. The nature and quantity of these nucleic acids and inhibitors in the respiratory tract vary from one person to another, and are totally unpredictable. Non-template nucleic acids can affect PCR amplification efficiency, the fluorescent signal growth curve and the cycle threshold (Ct) value in qPCR assays [19].

This paper introduces a protocol developed for conventional nested PCR amplification followed by sequencing of a 398-base cDNA amplicon of the nucleocapsid (N) protein gene to be used as a method for the detection of SARS-CoV-2 RNA, using cellular components of the specimens as the test material to raise the sensitivity of detection. The inhibitors carried over from the test material are greatly diluted in a nested PCR setting. Routine sequencing of the unique 398-bp nested PCR amplicon guarantees no-false positive test results. DNA sequencing of the PCR amplicon of the genomic nucleic acid of the pathogen is a well-established molecular test for detection of infectious agents which are difficult to culture [20].

2. Results

2.1. Single target sequencing for molecular diagnosis of SARS-CoV-2

An in-depth study of the region of the SARS-CoV-2 nucleocapsid (N) gene which the CDC selected to design its N1, N2 and N3 probes revealed that there is a highly conserved ~400-base segment of sequence with 5 single nucleotide polymorphisms between the N2 and N3 probes. Selection of a 398-base segment from this region for PCR detection and to prepare the template for DNA sequencing would be able to satisfy the FDA requirement for molecular diagnosis of viral infection [16]. The positional relationship between this 398-base sequence and those of the N1, N2 and N3 probes is illustrated in Figure 1.

Figure 1 A segment of the N gene sequence of SARS-CoV-2 retrieved from the GenBank database, sequence ID# LC528233, highlighted to show the sequences and positions of the 3 probes N1, N2 and N3 (in red) with their respective flanking primers (highlighted in yellow) and a 398-base segment (in bold and underlined; position: 28728-29125) with 5 single nucleotide polymorphisms (SNPs) highlighted in green. The sequence of the N gene targeted for SARS-CoV-2 RT-qPCR detection in China [21] is shaded in gray

```
28261 catctaaacg aacaaactaa aatgtctgat aatggacccc aaaatcagcg aaatgcaccc N1
28321 cgcattacgt ttggtggacc ctcagattca actggcagta accagaatgg agaacgcagt
28381 ggggcgcgat caaaacaacg tcggccccaa ggtttaccca ataatactgc gtcttggttc
28441 accgctctca ctcaacatgg caaggaagac cttaaattcc ctcgaggaca aggcgttcca
28501 attaacacca atagcagtcc agatgaccaa attggctact accgaagagc taccagacga
28561 attcgtggtg gtgacggtaa aatgaaagat ctcagtccaa gatggtattt ctactaccta
28621 ggaactgggc cagaagctgg acttccctat ggtgctaaca aagacggcat catatgggtt
28681 gcaactgagg gagccttgaa tacaccaaaa gatcacattg gcacccgcaa tcctgctaac N3
28741 aatgctgcaa tcgtgctaca acttcctcaa ggaacaacat tgccaaaagg cttctacgca
28801 gaagggagca gaggcggcag tcaagcctct tctcgttcct catcacgtag tcgcaacagt
28861 tcaagaaatt caactccagg cagcagtagg ggaacttctc ctgctagaat ggctggcaat
28921 ggcggtgatg ctgctcttgc tttgctgctg cttgacagat tgaaccagct tgagagcaaa
28981 atgtctggta aaggccaaca acaacaaggc caaactgtca ctaagaaatc tgctgctgag
29041 gcttctaaga agcctcggca aaaacgtact gccactaaag catacaatgt aacacaagct
29101 ttcggcagac gtggtccaga acaaacccaa ggaaattttg gggaccagga actaatcaga
29161 caaggaactg attacaaaca ttggccgcaa attgcacaat ttgcccccag cgcttcagcg N2
29221 ttcttcggaa tgtcgcgcat tggcatggaa gtcacacctt cgggaacgtg gttgacctac
```

The 398-base prototype sequence underlined is identical to the corresponding sequence in the N gene of the strain first isolated in Wuhan in December 2019 — GenBank Sequence ID: NC_045512.2. This version of sequence is present in most strains of SARS-CoV-2 subsequently isolated worldwide. However, strains with the 5 single nucleotide polymorphisms (highlighted in green) were first isolated in the United States. The dates and locations of the first sample collection for these 5 strains are summarized as follows, based on information retrieved from the GenBank database.

28800 A-to-T. Nasopharyngeal swab, CA, USA, 01-23-2020. Sequence ID: MN994467
28829 C-to-A. Nasal swab, NY, USA, 03-17-2020. Sequence ID: MT370913
28862 C-to-T. Oropharyngeal swab, MA, USA, 01-29-2020. Sequence ID: MT039888
28886 G-to-A. Nasopharyngeal swab, CA, USA, 02-06-2020. Sequence ID: MT106052
29103 C-to-T. Sputum of patient, TX, USA, 02-11-2020. Sequence ID: MT106054

Based on information retrieved from the GenBank database, this 398-base N gene nucleic acid sequence is specific for SARS-CoV-2. Except in the receptor-binding domain (RBD) where there is a 97.4% amino acid similarity between Guangdong pangolin coronaviruses and SARS-CoV-2, bat coronavirus RaTG13 is most closely related to SARS-CoV-2 in the remainder of the viral genome [22]. As illustrated below, alignment of the 398-base nucleic acid segment of the N gene of bat coronavirus RaTG13 (GenBank: MN996532, in red) against that of SARS-CoV-2 (in black) shows a 96.0% (1-16/398) similarity (the 16 nucleotide dissimilarities are highlighted green).

```
CAATCCTGCTAACAATGCTGCAATCGTGCTACAACTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAG
CAATCCTGCTAACAATGCTGCAATCGTGCTACAACTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAG
GTGGCAGTCAAGCTTCTTCTCGCTCTTCATCACGTAGTCGCAACAGTTCAAGAAACTCAACTCCAGGCAGCAGTAGGGGAACTTCC
GCGGCAGTCAAGCCTCTTCTCGTTCCTCATCACGTAGTCGCAACAGTTCAAGAAATTCAACTCCAGGCAGCAGTAGGGGAACTTCT
CCTGCTAGGATGGCTGGCAATGGCAGTGATGCTGCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTC
CCTGCTAGAATGGCTGGCAATGGCGGTGATGCTGCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTC
TGGTAAAGGCCAACAACAACAGGCCAAACTGTCACTAAGAAATCTGCTGCAGAGGCTTCTAAGAAGCCTCGGCAAAAACGTACTG
TGGTAAAGGCCAACAACAACAAGGCCAAACTGTCACTAAGAAATCTGCTGCTGAGGCTTCTAAGAAGCCTCGGCAAAAACGTACTG
CCACCAAACAATACAATGTAACACAAGCTTTTGGCAGACGTGGTCCAGAACAAA
CCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCAGAACAAA
```

The pair of PCR primers (underlined) defining this 398-base segment of amplicon are highly conserved among all known strains of SARS-CoV-2 and have not undergone mutations, based on search of the GenBank database. A single base mutation within a 21-base primer usually does not affect the result of heminested PCR unless the mutation occurs at the position binding the 3′ terminus of the primer.

SARS-CoV-1 gene (GenBank: AY274119.3) is not amplified by the primary PCR primer pair used in this protocol and there is < 90% similarity between the 398-base N gene sequence of the SARS-CoV-1 genome and that of the SARS-CoV-2 genome. Molecular misdiagnosis of SARS-CoV-2 based on nucleotide sequencing of this segment of the N gene is extremely unlikely.

2.2. Nested RT-PCR of cell lysates is needed to raise sensitivity for SARS-CoV-2 detection

In PCR research, a theoretically 100% efficient exponential amplification of the target DNA during repeated thermal cycling cannot be achieved even under strictly controlled experimental conditions. In the hands of most researchers, a properly designed PCR in the absence of any identifiable interfering substances in the sample matrix may amplify target DNA with a 90% efficiency [23, 24]. In molecular diagnosis of infectious diseases, there is always a possibility of false-negative results even when target DNA molecules are plentiful in the patient samples. After DNA purification steps are used before PCR, a 14% false-negative rate has been observed for hepatitis B virus detection, most probably due to incomplete removal of PCR inhibitors [25].

A high template/non-template nucleic acid ratio is one of the determinant factors in successful real-time PCR assays [19]. For example, persistent human papillomavirus (HPV) infection is the key factor in cervical carcinogenesis. During the long course of cancerous development, the viral load decreases from several thousand viral copies to about double digit per infected cell [15]. Virologic DNA real-time PCR assay is reported to be effective in identifying cervicovaginal cytology samples containing low-grade lesions with koilocytes which can have thousands of copies of human papillomavirus (HPV) per cell with a sensitivity rate of 95.5%. But the same real-time PCR test kit was only 58.26% effective in detecting high-grade CIN 3 lesions in which the viral load can be as low as 10 copies per cell [15, 26]. Non-template nucleic acids are potent PCR inhibitors which can greatly raise the Ct values of a qPCR assay when a non-template/template ratio is high in the reaction mixture.

False-negative test results for SARS-CoV-2 are expected in any PCR assays, including RT-qPCR when the viral load is low in the sample being tested. In testing low viral load samples, the laboratories must use highly sensitive assays to increase the detection sensitivity without causing false positives.

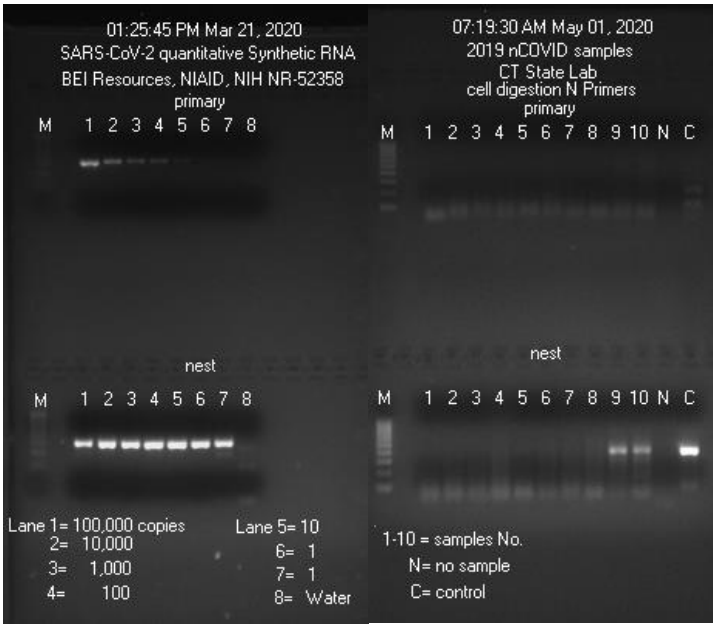
All diagnostic PCR assays depend on fluorescence intensity for evaluation. In the CDC RT-qPCR Assay Panel, the intensity of fluorescence emitted by a dye is measured by fluorometer, an instrument. In conventional PCR, the intensity of fluorescence emitted by a DNA/ethidium bromide complex is observed and evaluated with human eyes. A fluorometer used in qPCR is more sensitive than human eyes in measuring fluorescent signals, but it does not measure fluorescence emitted from a DNA complex. There are no products generated by the CDC RT-qPCR assays for further analysis whereas an amplicon of conventional PCR is composed of a mass of DNA molecules. The heminested PCR amplicon can be used as the template for DNA sequencing validation. Therefore, conventional PCR followed by DNA sequencing is the more appropriate testing platform for SARS-CoV-2 if a reliable test result is desired for patient management.

Nested, or two-round PCR, is a proven technique to mitigate the low sensitivity of conventional PCR in molecular diagnosis of infectious diseases, such as HPV infections [27] and borrelial spirochetemia [28]. This technique can be used for SARS-CoV-2 tests to increase detection sensitivity, as shown in Figure 2.

The left panel of Figure 2 is an image of agarose gel electrophoresis of the products of primary RT-PCR (upper half) and heminested PCR (lower half) showing that heminested PCR increased the sensitivity of one-round RT-PCR by ~1,000 fold in detecting a 398-base N gene segment of synthetic SARS-CoV-2 RNA. The number of copies of synthetic viral RNA added to each 25 μL primary RT-PCR mixture was calculated based on the analysis data supplied by BEI Resources, NIAD, NIH, Cat. No. NR-52358. As demonstrated, a single copy of viral RNA was detected with a robust amplicon band observed in lane 6 and lane 7 (lower half). In contrast, it needed 100-1,000 copies of viral RNA to generate a weak band in primary PCR, barely visible in lane 4 and lane 3 (upper half). The primary and heminested PCR primers and the PCR conditions are detailed in the Methods section.

The right panel of Figure 2 is an image of agarose gel electrophoresis showing two (2) positive heminested PCR products in lanes 9 and 10. There are no visible primary PCR products in any lanes. These 10 "negative" samples, each consisting of 0.5-1 mL of fluid (See Materials and Methods), were prepared and used as reference material by the Connecticut State Department of Public Health, Microbiology Laboratory Division on April 30, 2020 to assist local laboratories for developing their nucleic acid-based tests for SARS-CoV-2 under the CLIA program. Although these 10 samples were classified as negative by RT-qPCR, no Ct values were given. The cell lysates were extracted by phenol/chloroform and the RNAs were precipitated in ethanol as the test material (See Methods)

Figure 2 Agarose gel electrophoresis: Heminested PCR raises the sensitivity for detection of a 398-bp N gene segment in synthetic SARS-CoV-2 RNA (Left) and of viral RNA extracted from infected human cells in respiratory tract specimens (Right: Amplicons of cDNA of viral RNA detected in lanes 9 and 10 from reference samples #1-10 classified as negative by RT-qPCR assays)



The laboratory data presented in Figure 2 show that two-round PCR is needed for the detection of amplicons when this technology is used for molecular identification of SARS-CoV-2 in respiratory tract samples. The material to be tested should include cellular components of the specimen for maximizing the sensitivity of detection.

2.3. False positives of RT-qPCR assay for SARS-CoV-2

PCR, including qPCR, is known to generate both false-negative and false-positive results when this DNA replication tool is used for the diagnosis of infectious diseases.

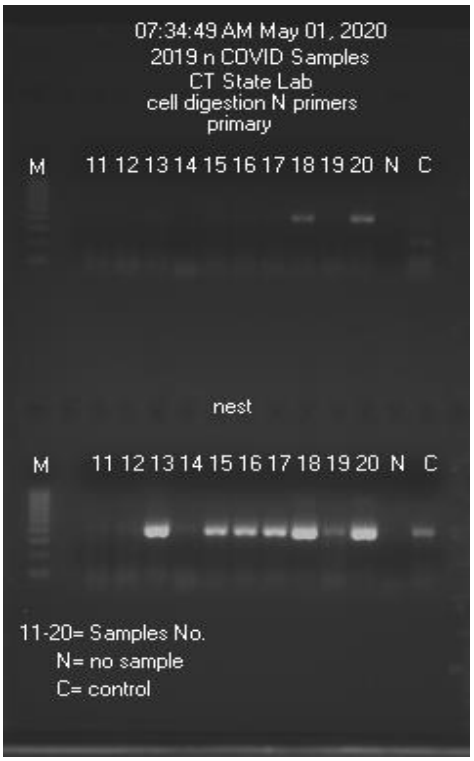
In an international external quality assessment proficiency testing program for MERS-CoV, a Coronavirus, laboratories using RT-qPCR assay on simulated samples with viral load equivalent to throat swab MERS-CoV RNA concentrations generated an 84.6% detection sensitivity and 8.1% false-positive results [29]. False-negative and false-positive rates are usually much higher in routine testing than those observed in regulatory proficiency test programs.

It is not unexpected to find false-positive results generated by RT-qPCR when this tool is used to test for SARS-CoV-2.

On April 30, 2020, in addition to the 10 negative samples referenced above the Connecticut State Department of Public Health, Microbiology Laboratory Division also prepared 10 reference samples, which were diagnosed as positive for SARS-CoV-2 by the CDC RT-qPCR assays, in support of the local laboratories to develop their tests for the virus.

According to the documents accompanying these 10 positive samples, they all generated an N1 Ct value between 16 and 29.78 and an N2 Ct value between 15.05 and 29.4 except one with an N2 Ct 67.86. The R. P. (RNase P gene) Ct values for these 10 samples ranged from 3.31 to 25.13. Based on these Ct values, all samples were interpreted as confirmed positive and used as standard reference material for developing SARS-CoV-2 assays. However, only 7 of these 10 samples were found to be positive by nested PCR amplification (Figure 3).

Figure 3 Agarose gel electrophoresis showing that only 7 of the 10 samples classified as positive for SARS-CoV-2 by RT-qPCR could be confirmed by conventional nested PCR followed by DNA sequencing of a 398-bp cDNA amplicon of the N gene



As demonstrated in Figure 3, primary PCR products after amplification of 30 thermal cycles were only visualized on 2 of 7 validated positive samples (lanes 18 and 20). Five of the 7 positives (5/7) needed nested PCR for detection. All seven positive nested PCR products were proven to consist of a 398-base segment of SARS-CoV-2 N gene by DNA sequencing. Six (6/7) were of the protype (See Section *2.4.* Figures 4 and 5). Sample #13 belongs to a newly discovered mutant first isolated from a human specimen collected in the State of New York on March 17, 2020 (See Section *2.4.* Figures 6 and 7).

2.4. Routine sequencing for PCR product validation and single nucleotide polymorphisms

For emerging serious virus infectious diseases, such as Ebola, the European CDC's laboratory criteria of case definition are 1) Detection of virus nucleic acid in a clinical specimen and confirmation by sequencing or a second assay on different genomic targets; or 2) Isolation of Ebola virus from a clinical specimen [30].

A recent "Report from the American Society for Microbiology COVID-19 International Summit" on the Value of Diagnostic Testing for SARS–CoV-2/COVID-19 pointed out that one of the new challenges in the current molecular diagnostics is to optimize the current multiple targets to a single target [31].

By following these recommendations and the FDA's guidance for molecular diagnosis of another viral infection [16], routine bi-directional sequencing of a >100-bp PCR amplicon of the viral genome was designed as the proper molecular testing platform for SARS–CoV-2. Examples are presented by analysis of the computer-generated 398-base N gene sequences on the reference samples supplied by the Connecticut State Department of Public Health Microbiology Laboratory Division as follows.

Figure 4 Electropherogram showing DNA sequence of the nested PCR amplicon visualized in Lane 20 of Figure 3. Forward Co4 PCR primer was the sequencing primer. Note: the base C in position 14 is underlined

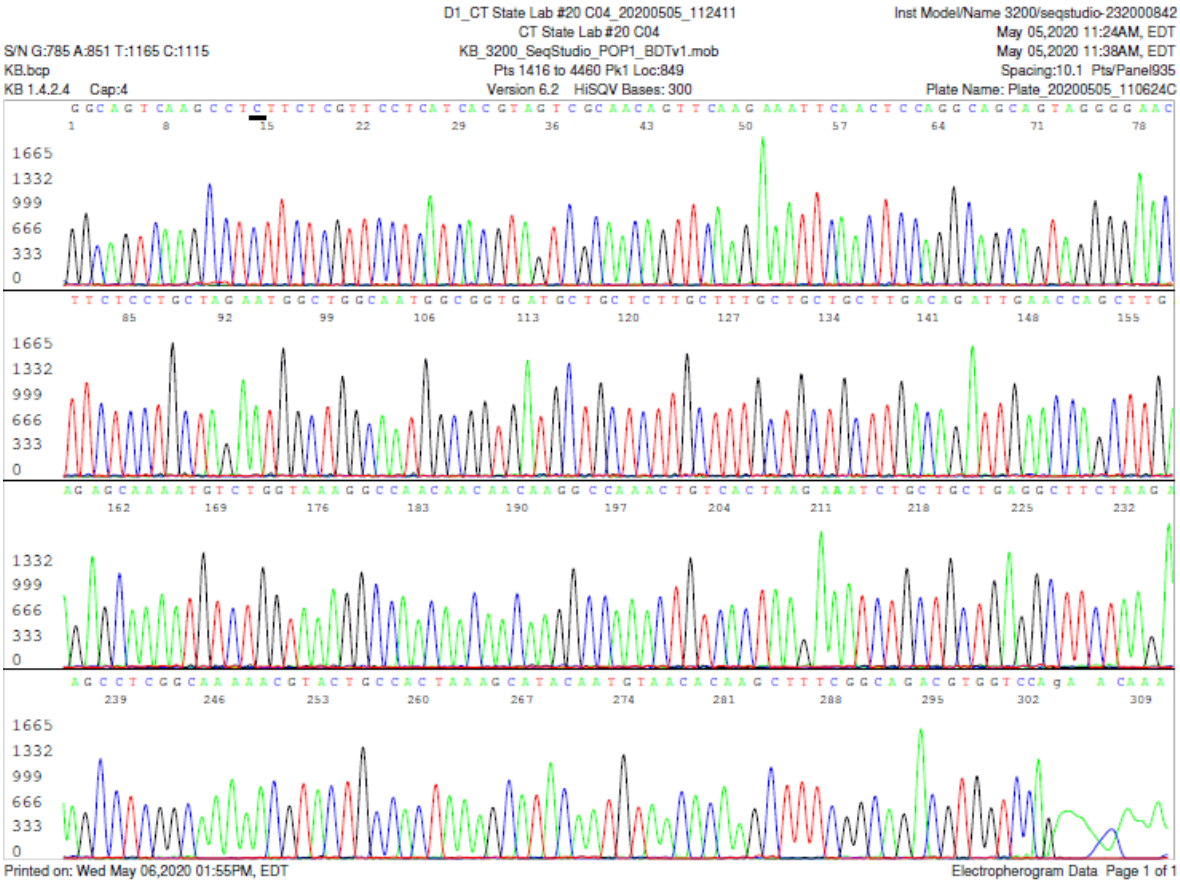
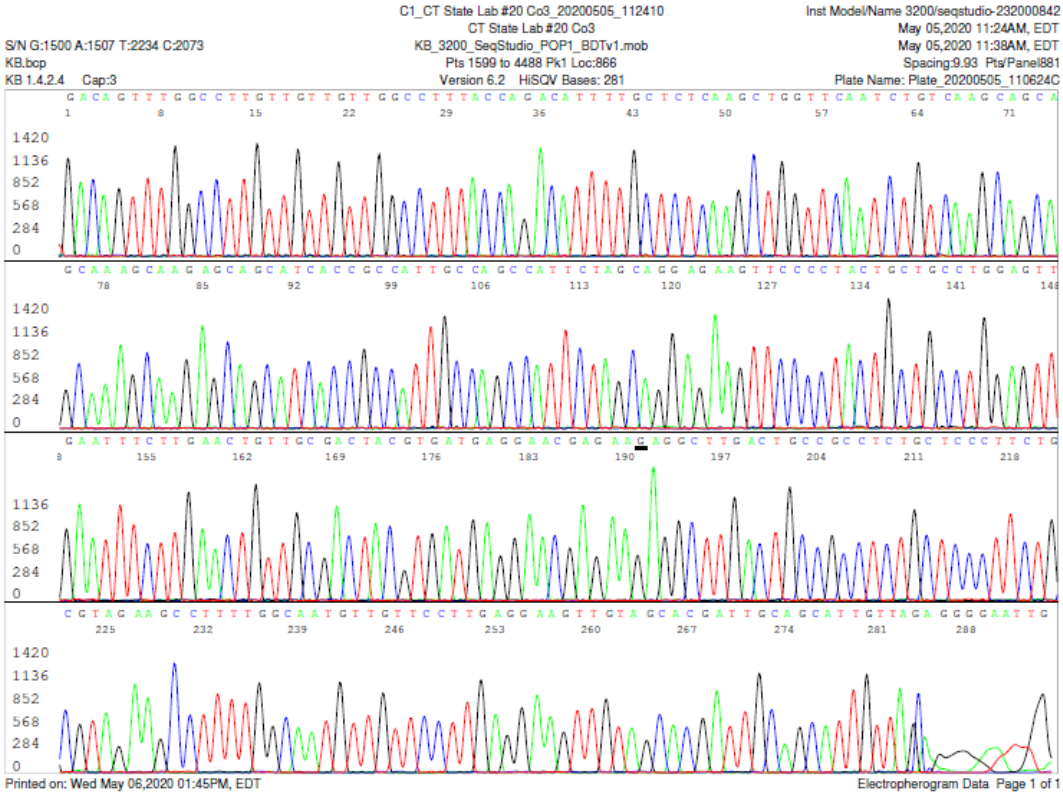


Figure 5 Electropherogram showing DNA sequence of the nested PCR amplicon visualized in Lane 20 of Figure 3. Reverse Co3 PCR primer was the sequencing primer. Note: the base G in position 191 is underlined



Connecting the two sequences illustrated in Figures 4 and 5 after all the complementary bases were converted to those for a 5′–3′ reading resulted in a composite sequence fully matching a 398-base sequence illustrated in Figure 1, position 28728-29125 which is generally accepted as the protype sequence for SARS-CoV-2.

Figure 6 Electropherogram showing DNA sequence of the nested PCR amplicon visualized in Lane 13 of Figure 3. Forward Co4 PCR primer was the sequencing primer. Note: the base A in position 14 is underlined

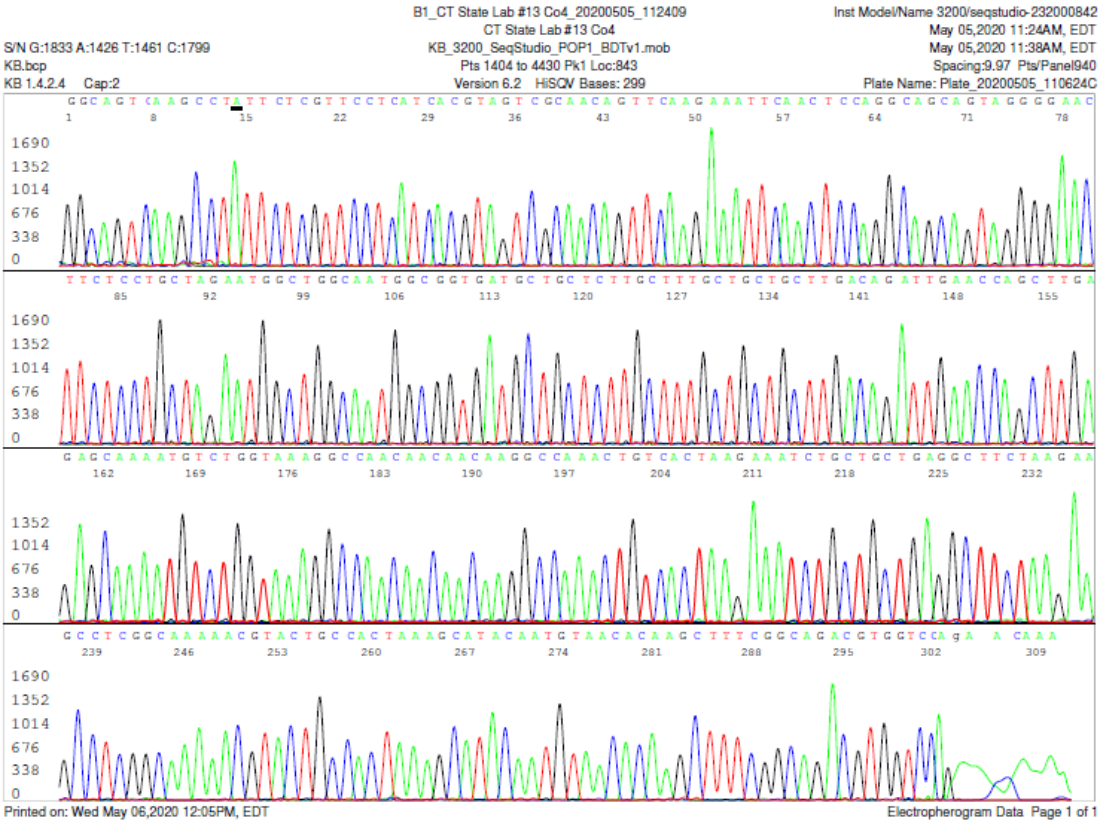
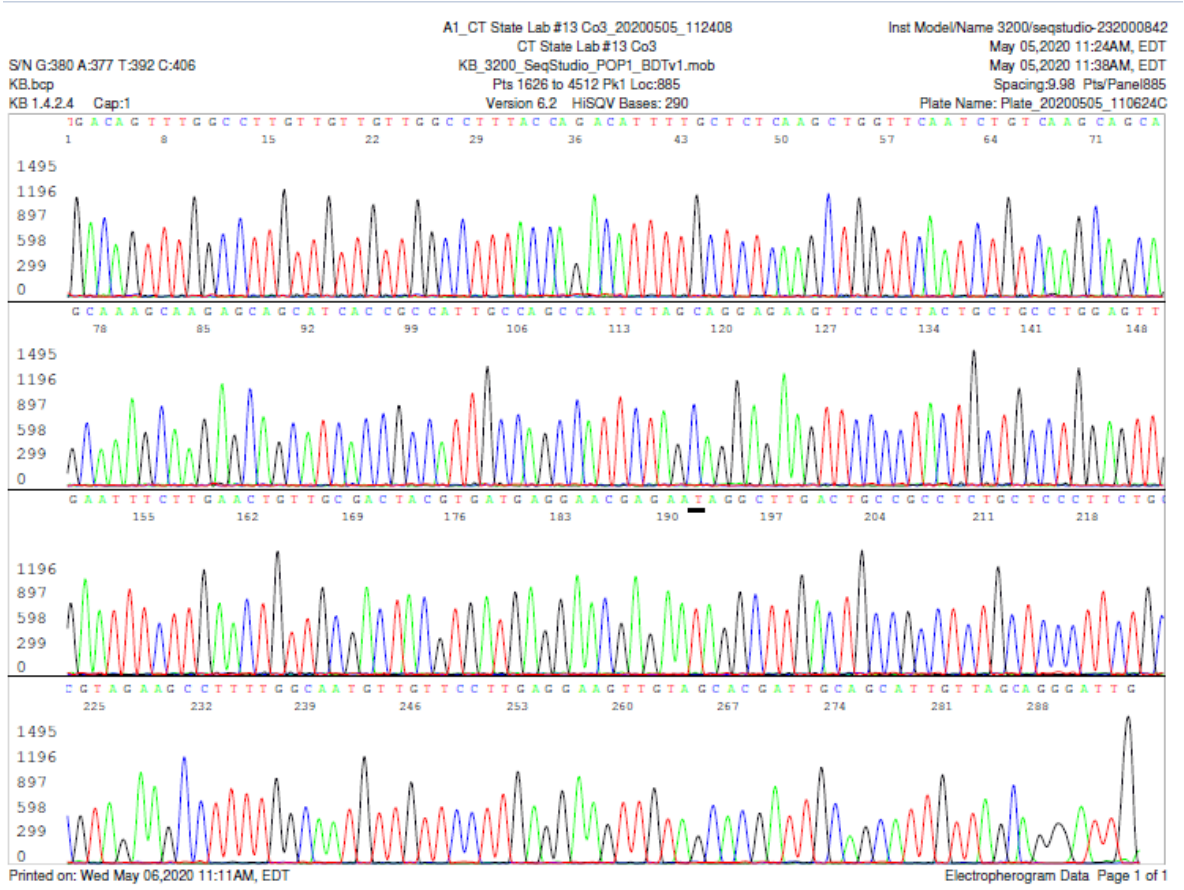
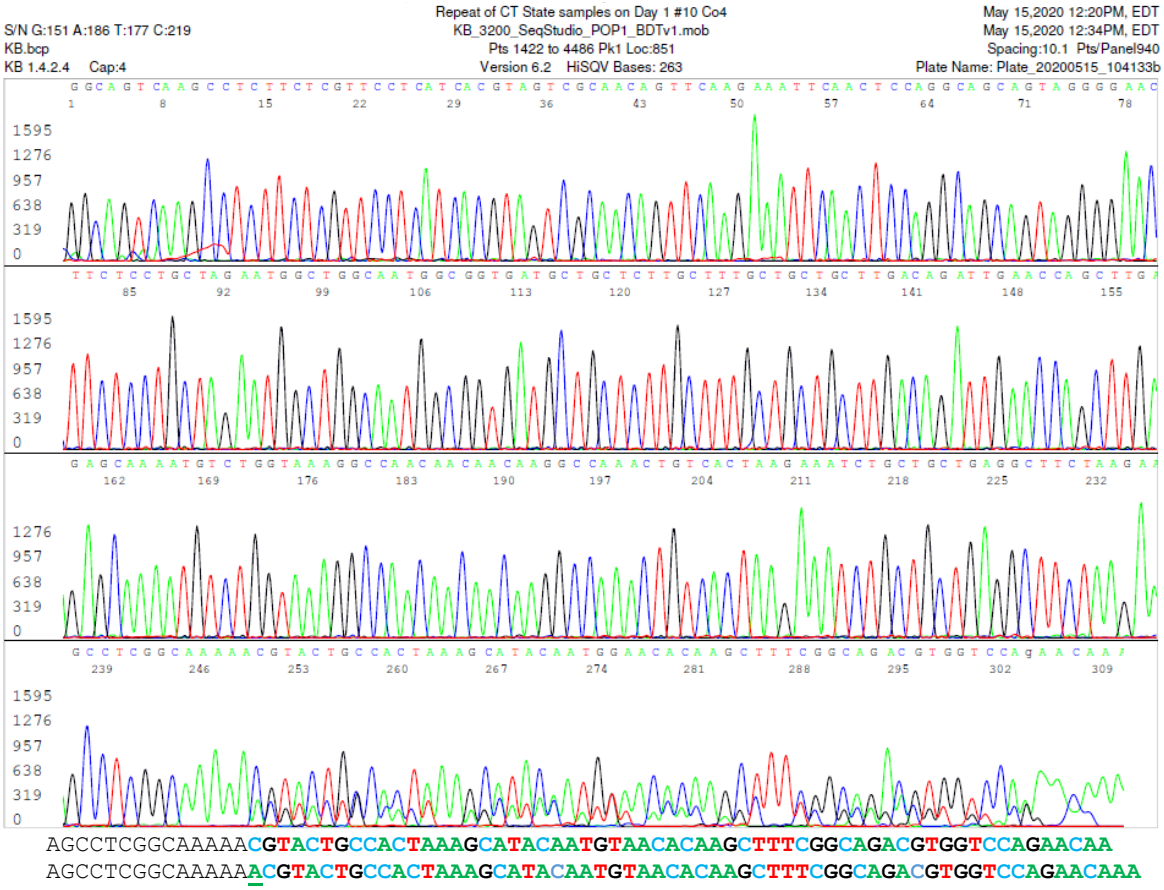


Figure 7 Electropherogram showing DNA sequence of the nested PCR amplicon visualized in Lane 13 of Figure 3. Reverse Co3 PCR primer was the sequencing primer. Note: the base T in position 191 is underlined



Connecting the two sequences illustrated in Figures 6 and 7 after all the complementary bases were converted to those for a 5′–3′ reading resulted in a composite sequence fully matching a 398-base sequence illustrated in Figure 1, position 28728-29125 except for a single base mutation of C-to-A in position 28829, as annotated in GenBank Sequence ID: MT370913.

Figure 8 Electropherogram showing two superimposed SARS-CoV-2 N gene sequences in the nested PCR amplicon visualized in Lane 10 of Figure 2, right panel. Forward Co4 PCR primer was the sequencing primer. One sequence was from a novel mutant with a single nucleotide A insertion at position 250, and the other was from a wildtype parental virus



The computer-generated electropherogram illustrated in Figure 8 shows one protype N gene sequence of SARS-CoV-2 from position 1 to position 249. But there was a single nucleotide A insertion at position 250 in the second N gene, creating a mutant while the wildtype parental virus was infecting the human host. The chance for one patient being infected by two SARS-CoV-2 strains from the very beginning is extremely low. A virus mutant with a nucleotide A insertion in this segment of the N gene has not been reported in the literature nor deposited into the GenBank. In Figure 8, the extra nucleotide in the mutant gene caused a frameshift in DNA sequencing after the nucleotide A insertion. The two components of each double peak in different colors are aligned against each other in sequence and presented immediately below the electropherogram. The inserted base A is underlined.

This sequencing pattern showing an "A" insertion in one of the two templates at position 250 was reproduced for 3 times to rule out possible sequencing artifacts. Sequencing from the opposing direction on this amplicon composed of two homeologous DNA templates failed to generate a readable sequence due to insertion of the "A" nucleotide into the mutant gene too close to the site of the reverse Co3 sequencing primer.

In summary, the results of re-testing the cellular components of 20 reference samples of nasopharyngeal and oropharyngeal swab rinses by heminested RT-PCR amplification followed by

nucleotide sequencing showed that SARS-CoV-2 was not found in 3 of the 10 (3/10) reference samples classified as positive by RT-qPCR, and that 2 of the 10 (2/10) reference samples classified as negative by RT-qPCR in fact contained SARS-CoV-2.

Among the 9 positive samples, one isolate showed a C-to-A single nucleotide mutation in the 398-base segment of the N gene targeted for amplification. According to information retrieved from the GenBank database, this mutation was discovered first in a specimen collected in New York State on March 17, 2020. And only 5 cases with this mutation were reported to the GenBank as of May 13, 2020.

In addition, one of the two positive isolates from the 10 RT-qPCR-negative samples in fact contained two viral genomes, one from a mutant with a single nucleotide insertion in the N gene which was not found in the GenBank database, and one from a wildtype parental virus. Co-infection of one patient by two strains of SARS-CoV-2, a mutant and the wildtype parental virus, has not been reported and may pose a challenging problem in molecular diagnosis of SARS-CoV-2

2.5. Unintended primer extension in RT-qPCR

The CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel [11] advises to use cell-free samples prepared by the QIAamp Viral RNA Mini Kit as the material to be tested. But it also requires proof of human genomic nucleic acids in the same samples for a valid negative SARS-Cov-2 RNA result. As proof of the presence of sufficient human cell nucleic acids in the material being tested, the CDC protocol sets a cycle threshold (Ct) value within 40.00 cycles (< 40.00 Ct) in the human RNase P gene RT-qPCR channel for positive cut-off.

The sequences of the primers and probe for the RNase P gene RT-qPCR as an Internal Process Control (IPC) are given as follows [32] (color-highlighted for discussion convenience).

RNAse P Forward Primer 5′-AGA TTT GGA CCT GCG AGC G-3′

RNAse P Reverse Primer 5′-GAG CGG CTG TCT CCA CAA GT-3′

RNAse P Probe 5′-FAM – TTC TGA CCT GAA GGC TCT GCG CG – BHQ-1-3′

These primers and probe were designed to amplify a 65-base segment of *Homo sapiens* ribonuclease P/MRP subunit p30 (RPP30). Sequence ID: NM_001104546. The relevant sequence in this segment of the gene is retrieved from the GenBank database and pasted below with primer-matching colors to indicate the positions of the CDC-recommended primers and the probe which is typed in red in reverse complement (Figure 9).

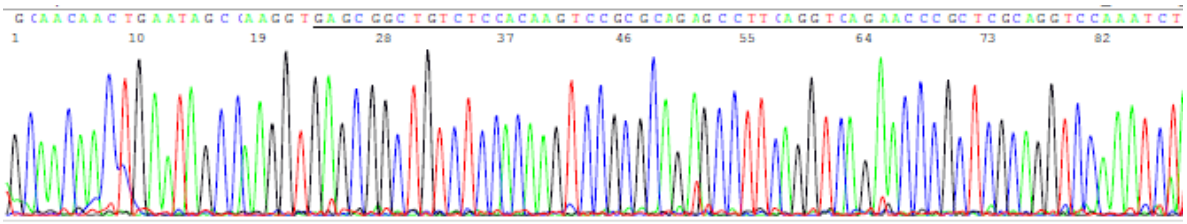
Figure 9 Segment of DNA sequence retrieved from the GenBank database (NM_001104546) showing the color-highlighted positions of primers and probe used in the CDC RT-qPCR assay kit and the sequence of the 160-bp (underlined) amplicon used as the template to generate the sequence illustrated in Figure 10

```
240  CTAGTTAAAATTTTAATTGGTCTTGATTTTCCCTGTACAATTGGCAAAGTTGTGAAGAGT  181
180  TCAGAAACAGCTACTGGTTTTTCAATTTCCTGTTTCTTTTCCTTAAAGTCAACGATATGA  121
120  TTGATAGCAACAACTGAATAGCCAAGGTGAGCGGCTGTCTCCACAAGTCCGCGCAGAGCC  61
60   TTCAGGTCAGAACCCGCTCGCAGGTCCAAATCTGCAAACACCGCCATGCTGAAGTCCCAT  1
```

It was assumed that a 65-base segment of the RPP30 gene defined by the two primers would be amplified as shown in Figure 9 in the RT-qPCR testing platform. However, it has never been

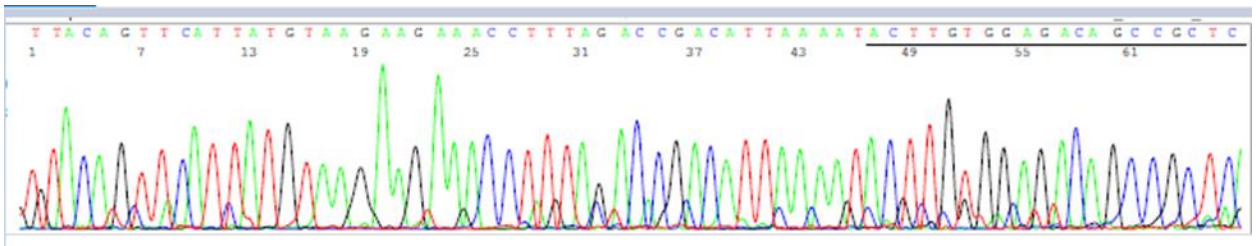
demonstrated that this sequence has actually been amplified in real practice because the inter-primer sequence is too short for the currently used sequencing technologies. To prove that this 65-base sequence is actually present in the clinical sample, a longer amplicon (160-base, underlined in Figure 9) is needed to be used as the sequencing template to generate a sequence illustrated in Figure 10.

Figure 10 Electropherogram of a 65-base sequence (underlined) presumed to be generated during RT-qPCR using the CDC SARS-CoV-2 test kit, but it has never been shown in real practice



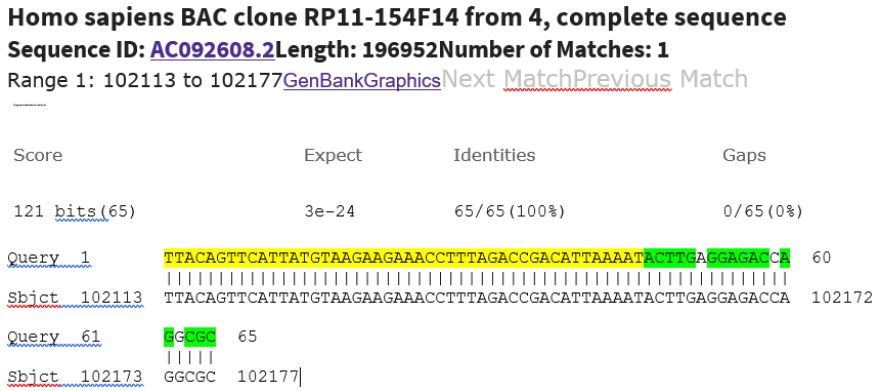
There is a single copy of RPP30 DNA sequence located at 10q23.31 locus on chromosome 10 in each haploid human genome and there are two copies in a diploid human cell [33]. RPP30 has been used as a reference for the autosomal chromosomes in research when there is abundant human chromosomal material in the sample [33, 34]. However, in cell-free fluids derived from respiratory tract specimens there are no intact human cells in the sample being tested. RPP30 gene is detected in the viral RNA extracts only when some broken human cell nuclear parts have not been completely removed from the cell-free samples. In the absence of a preferred template, the PCR primers may anneal to a non-target DNA with partially matching sequence and initiate an unintended primer extension or PCR. Such an example is illustrated in Figure 11.

Figure 11. Sequencing electropherogram showing PCR amplification of an unintended DNA segment by the CDC RNase P Reverse Primer (underlined)



The sequence illustrated in Figure 11 was generated during an attempt of using the CDC RNase P primers to induce a 65-bp PCR amplicon for further analysis as an internal control which can be validated by DNA sequencing. In this electropherogram, the binding site for the CDC RNase P reverse primer is underlined. Submission of this sequence to GenBank for BLAST algorithmic analysis led to a return illustrated in Figure 12 (yellow and green added by author).

Figure 12 In the absence of a fully matching template, the primers designed for the CDC RT-qPCR assay kit may anneal to partially matching DNA and initiate an unintended PCR



As shown in Figure 11, PCR amplification as often applied in clinical diagnostics is not absolutely specific. There are always some irrelevant products, visible or invisible, generated as a result of enzymatic DNA replications in addition to those represented in the main sequence. The main sequence in this electropherogram shows that the CDC RNase P reverse primer found and annealed to a segment of DNA which shares 16 (highlighted in green, Figure 12) of its 20 bases in sequence to initiate an unintended PCR (compare the green-highlighted sequence in Figure 12 with the primer sequence underlined in Figure 11). The sequence highlighted in yellow is totally different from that of the probe (Figure 9).

The findings described in this section suggest that non-specific hybridization or annealing between non-target DNA and the primer or probe used in the CDC RT-qPCR test kits for SARS-CoV-2 may have contributed to the causes for false-positive results especially when the Ct cut-off value has been set as high as 40.0 [11].

Other investigators also reported that the CDC N2 and N3 primer sets can amplify irrelevant nucleic acids in RT-qPCR and in conventional PCR [4].

In an FDA publication titled "Real-Time Fluorescent RT-PCR Kit for Detecting SARS-CoV-2", the Instructions for Use repeatedly emphasize that the Ct values for positive result on all controls and samples should be no higher than 37 [35]. Setting a 40.00 Ct value as the cut-off point for the presence of RNase P gene or for the presence of SARS-CoV-2 N gene in the sample being tested deviates from common practice and from the FDA guidance. Accepting a questionable Ct value as evidence for the presence of RNase P gene in the sample being tested may have allowed many invalid test results to be reported as negatives.

2.6. *Human BRCA1 gene as internal extraction control*

Since the segment of RNase P gene selected for the CDC RT-qPCR test panel was not always PCR-amplifiable for validation by conventional PCR followed by sequencing, a segment of human *BRCA*1 gene was selected as the internal cellular extraction control.

A pair of AG1 and R2 PCR primers (see Materials and Methods section) was designed to initiate a 409-bp *BRCA*1 gene primary PCR. The primary PCR product was re-amplified by a pair of AG2/R2 heminested PCR primers to generate a 323-bp amplicon which can be used as the template for sequencing validation. Demonstration of a segment of *BRCA*1 gene amplicon was accepted as reliable physical evidence to confirm that sufficient human nucleic acids had been extracted into the sample being tested in the SARS-CoV-2 negative cases. While the RNase P gene in the 20 reference samples supplied by the Connecticut State Department of Public Health Microbiology Laboratory Division could not be detected, the *BRCA* 1 gene was demonstrated in all these samples as shown in Figures 13 and 14.

Figure 13 Image of agarose gel electrophoresis showing a 323-bp heminested PCR amplicon of human *BRCA*1 gene in all 20 reference samples tested. Lane 11=water negative control. Lane 12= known human cell extract positive control

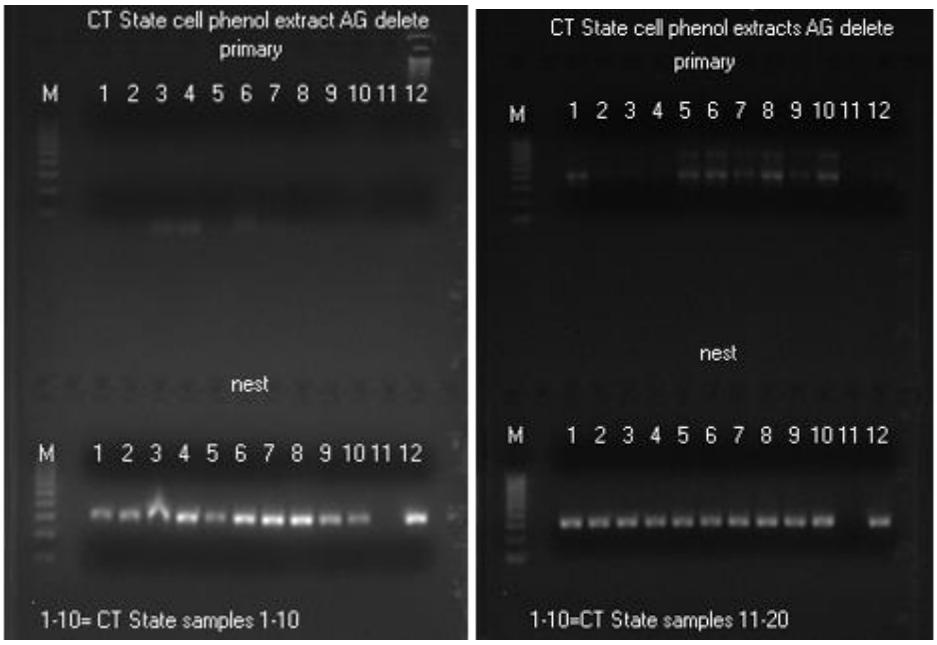
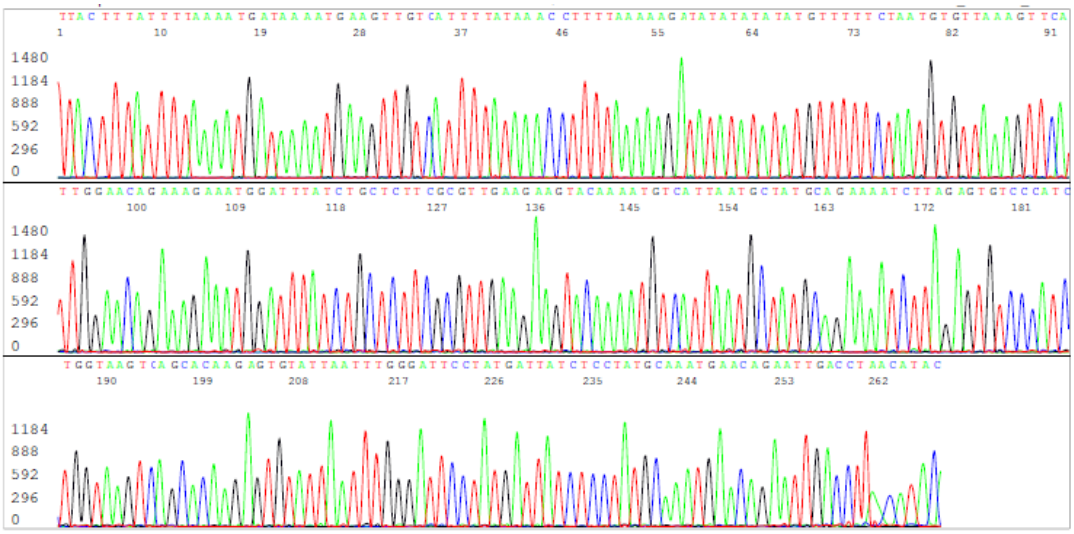


Figure 14 Sample of electropherograms confirming a segment of human *BRCA*1 gene sequence of the nested PCR amplicons shown in lower half of Figure 13. The 21 bases in the end are the binding site for R2 primer (5′-GTATGTAAGGTCAATTCTGTTC-3′)



3. Discussion

This article has demonstrated that for molecular diagnosis, testing the cellular components for SARS-CoV-2 rather than cell-free fluids only on respiratory tract specimens by conventional nested PCR and using the traditional phenol/chloroform method for RNA sample preparation can reduce the number of false-negatives. Using nucleotide sequencing to validate every positive PCR amplicon practically eliminates all false-positive results. For timely accurate molecular diagnosis and characterization of infectious agents, concerned scientists have been urging to bring microbial sequencing to the hospital laboratories to improve individual and population health [36, 37].

Nested PCR is a simple method to increase the detection sensitivity of the conventional PCR technology. Cross contamination which is often used as a reason to object using nested PCR in diagnostic laboratories can be mitigated by elimination of transferring PCR products by micropipettes and proper training of the testing staff. Cross contamination is a function of the laboratory performing PCR, not an inherent part of the nested PCR technology [28].

RT-qPCR tests are known to generate both false-negative and false-positive results [4, 6, 7, 9]. Highly sensitive and accurate laboratory tests for SARS-CoV-2 are needed for certain populations, especially for the people of advanced age living in long-term care facilities and their care takers. Patients with false-negative test results may transmit the virus to family, friends, or care-givers. Uninfected residents in long-term care facilities with false-positive results may be isolated in rooms with Covid-19 patients, which puts them at risk of becoming true positives. Convalescent hospitalized patients must be tested by an extremely sensitive, no false-positive nucleic acid test prior to being discharged into the communities.

Long-term care facilities with exceptionally high COVID-19 death tolls among their residents may consider re-testing the residues of the respiratory tract samples of the deceased collected and tested prior to their death. If false-negative and false-positive test results were found, the false test results might have contributed to spreading of SARS-CoV-2 in the institution. Then corrective measures can be made in the infection control practices to lower the rate of cross infection among the residents.

Highly sensitive, no-false positive SARS-CoV-2 tests are needed in hospitals with head-and-neck surgical department to screen patients before admission for protection of the medical staff involved [9].

Extremely accurate SARS-CoV-2 laboratory tests are especially important as businesses attempt to return to normal operation in order to avoid raising false alarms of a re-emerging outbreak. False-positive test reports can easily create unnecessary panic resulting in negative impacts on local economies.

The current SARS-CoV-2 nucleic acid test platforms overly depend on the supply of commercial RNA extraction devices, such as the QIAamp viral RNA preparation kits. However, better options are readily available. For example, the abilities of QIAamp blood kit (Qiagen, Inc., Hilden, Germany) and another commercial kit to extract the hepatitis B virus (HBV) DNA template from serum for amplification by PCR were evaluated and compared with that of the standard phenol-chloroform method. Differences in the sensitivities of the three methods were revealed by nested PCR of HBV DNA extracted from serially diluted hepatitis B e antigen (HBeAg)-positive (high-titer) serum. Phenol-chloroform was found to be a 10^6 times more sensitive extraction method than the QIAamp blood kit; and nested PCR was found to be 10^6 times more sensitive than one-round PCR in detection of HBV DNA in serum samples [25]. In nested PCR, the PCR inhibitors carried over from the sample into the primary PCR are further diluted by about 100-fold. The value of nested PCR followed by DNA sequencing for detection of SARS-CoV-2 was previously reported by Nao and colleagues [38].

As demonstrated in the Results section of this article, using the standard phenol-chloroform method to prepare viral RNA from cell lysates and nested RT-PCR amplification followed by DNA sequencing for SARS-CoV-2 detection, 2 false negatives and 3 false positives were found in 20 reference samples which had been tested by the CDC RT-qPCR kits.

In addition, one of the seven (1/7) RT-qPCR positive samples was found unexpectedly to contain an isolate which belongs to a strain of SARS-CoV-2 with single C-to-A mutation in the 398-bp N gene segment targeted for amplification. This newly discovered strain was first recognized in a specimen collected in the State of New York on March 17, 2020, and then found in specimens collected in the District of Columbia and Connecticut in the next two weeks, according to information retrieved from the GenBank database. It is not known how many of the Covid-19 patients were caused by this locally developed “mutant” in the tri-state area of New York, Connecticut and New Jersey.

Another surprise in testing the 20 reference samples prepared by the Connecticut State Department of Public Health Microbiology Laboratory Division was the finding of a SARS-CoV-2 mutant with a novel single nucleotide A insertion within the 398-base DNA segment sequenced. This novel mutant and the wildtype parental virus apparently co-infected the host cells collected for testing and their N genes were co-amplified in the nested PCR process. This novel viral mutant was present in a person with a false-negative test result and was capable of infecting the host cells as the parental virus. It is not known if such a mutant can be transmitted to another individual to cause a new infection as the wildtype parental virus can.

As of early May 2020, there were at least 198 recurrent mutations identified among global isolates of SARS-CoV-2 [39]. For patient management, it is neither practical nor necessary to routinely check all these known mutations in the genome of every isolate for diagnostic purpose.

However, it is feasible to routinely perform a bi-directional DNA sequencing of a 398-bp PCR amplicon of the N gene on every positive sample to categorize each isolate into one of the 6 SARS-CoV-2 clades which are known to circulate in the United States to track real-time movement of these viral strains in order to facilitate studies of their possible relevancy to transmissibility and pathogenicity.

The major limitation of this method is that the nested PCR and nucleotide sequencing technologies cannot be readily automated. The test takes 24 to 48 hours to generate a result because the procedures of cell digestion, two PCR amplifications and DNA sequencing take more than 10 hours of instrument time to complete. As for all PCR-based nucleic acid detection methods, unexpected mutation of the N gene in the SARS-CoV-2 genome affecting the binding site for the 3′ terminus of any of the PCR primers may cause false-negative results. If a patient does not have SARS-CoV-2 infected cells in the upper respiratory tract, or the specimen does not contain any virus-infected cells, this method may generate false-negative test results.

4. Materials and Methods

4.1. Test materials

The specimens used for this study were 10 negative and 10 positive reference samples which were specially prepared by the Connecticut State Department of Public Health, Microbiology Laboratory Division, on April 30, 2020 and packaged in a Styrofoam box containing dry ice to be used in support of the local diagnostic laboratories in the State of Connecticut to develop their laboratory tests for detecting SARS-CoV-2 virus in respiratory tract specimens. All samples were in 0.5-1 mL of saline as the media in plastic vials with screw caps and processed on the same day in Milford Molecular Diagnostics Laboratory, in Milford, Connecticut.

According to the Connecticut State Department of Public Health, Microbiology Laboratory Division, these samples were all clinical specimens, namely nasopharyngeal or oropharyngeal swabs placed in saline, that were received at the Connecticut State Department of Public Health, Microbiology Laboratory Division for the purpose of testing for SARS-CoV-2. No culture, spiking/or diluents were added or performed on these specimens. These 20 samples were previously tested by the CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel as negative for SARS-CoV-2 or positive for SARS-CoV-2, and were used as reference samples for evaluation of other laboratory-developed assays for SARS-CoV-2 in nasopharyngeal and oropharyngeal swab specimens.

4.2. RNA extraction

The entire content of each sample was transferred to a graduated 1.5 mL microcentrifuge tube and centrifuged at ~16,000× g for 5 min to pellet all cells and cellular debris. The supernatant was discarded except the last 0.2 mL which was left in the test tube with the pellet.

To each test tube containing the pellet, 200 μL of digestion buffer containing 1% sodium dodecyl sulfate, 20mM Tris-HCl (pH 7.6), 0.2M NaCl and 700 μg/mL proteinase K, modified from a protocol for releasing cellular RNase P genes from HeLa cells [40], was added. The mixture was digested at 47°C for 1 hr in a shaker.

An equal volume (400 μL) of acidified 125:24:1 phenol:chloroform:isoamyl alcohol mixture (Thermo Fisher Scientific Inc.) was added to each tube. After vortexing for extraction and centrifugation at ~16,000×

g for 5 min to separate the phases, 200 μL of the aqueous supernatant without any material at the interface was transferred to a new 1.5 mL microcentrifuge tube.

To the 200 μL of phenol/chloroform-extracted aqueous solution, 20 μL of 3M sodium acetate (pH 5.2) and 570 μL of 95% ethanol were added. The mixture was placed into a cold metal block in a -15 to -20°C freezer for 20 min. The precipitated nucleic acids were centrifuged at ~16,000× g for 5 min and washed with 700 μL of cold 70% ethanol.

After a final centrifugation at ~16,000× g for 5 min, the 70% ethanol was completely removed by a fine-tip pipette, and the microcentrifuge tubes with opened caps were put into a vacuum chamber for 10 minutes to evaporate the residual ethanol. The nucleic acid in each tube was dissolved in 50 μL of M.B. grade water. All samples were tested immediately.

As positive control, a 0.2 mL aliquot of suspension of nasopharyngeal and oropharyngeal cells previously collected from normal individuals, tested negative for SARS-CoV-2, and stored in normal saline at -20°C was spiked with 400 copies of Quantitative Synthetic RNA from SARS-Related Coronavirus 2, Cat. No. NR-52358 supplied by BEI Resources, NIAD, NIH. This positive control tube was processed along with the reference samples for digestion and extraction.

The negative control consisted of 0.2 mL aliquot of suspension of nasopharyngeal and oropharyngeal cells previously collected from normal individuals, tested negative for SARS-CoV-2, and stored in normal saline at -20°C.

4.3. PCR primers

Based on the sequence retrieved from the GenBank database and presented in Figure 1, 3 primers were designed for primary and heminested PCR amplification of a 398-bp cDNA fragment to be used as the template for nucleotide sequencing confirmation of nucleic acid from the SARS-CoV-2. Their sequences and positions are listed below (see Figure 1).

Primary PCR primers
Forward primer Co1 5′-ACATTGGCACCCGCAATCCTG-3′ 28715-28735
Reverse primer Co3 5′-TTTGTTCTGGACCACGTCTGC-3′ 29105-29125
Heminested PCR primers
Forward primer Co4 5′-CAATCCTGCTAACAATGCTGC-3′ 28728-28748
Reverse primer Co3 5′-TTTGTTCTGGACCACGTCTGC-3′ 29105-29125

The Sequences of the primary and heminested PCR primers for the *BRCA*1 gene are listed as follows.

Primary PCR primers for a 409-bp *BRCA*1 gene amplicon are:
Forward AG1 primer 5′-AAGGGGTTGGCAGCAATATGTG-3′
Reverse R2 primer 5′-GTATGTAAGGTCAATTCTGTTC-3′

Heminested PCR primers for a 323-bp *BRCA*1 gene amplicon are:
Forward AG2 primer 5′- GACGTTGTCATTAGTTCTTTGG-3′ (also for sequencing)
Reverse R2 primer 5′-GTATGTAAGGTCAATTCTGTTC-3′

4.4. Primary RT-PCR conditions for SARS-CoV-2 gene amplification

To initiate the primary RT-PCR, a total volume of 25 μL mixture was made in a PCR tube containing 20 μL of ready-to-use LoTemp® PCR mix with denaturing chemicals (HiFi DNA Tech, LLC, Trumbull, CT, USA), 1 μL (200 units) of Invitrogen SuperScript III Reverse Transcriptase, 1 μL (40 units) of Ambion™ RNase Inhibitor, 0.1 μL of Invitrogen 1 M DTT (dithiothreitol), 1 μL of 10 μmolar Co1 forward primer in TE buffer, 1 μL of 10 μmolar Co3 reverse primer in TE buffer and 1 μL of sample RNA extract, synthetic N gene RNA extract as positive control or water as negative control.

The ramp rate of the thermal cycler was set to 0.9 °C/s. The program for the temperature steps was set as: 47°C for 30 min to generate the cDNA, 85°C 1 cycle for 10 min, followed by 30 cycles of 85°C 30 sec for denaturing, 50°C 30 sec for annealing, 65°C 1 min for primer extension, and final extension 65°C for 10 minutes.

4.5. Heminested PCR conditions for SARS-CoV-2 gene amplification

The heminested PCR mixture was a 25 μL volume of complete PCR mixture containing 20 μL of ready-to-use LoTemp® mix, 1 μL of 10 μmolar Co4 forward primer, 1 μL of 10 μmolar Co3 reverse primer and 3 μL of water.

To initiate the heminested PCR, a trace (about 0.2 μL) of primary PCR products was transferred by a micro-glass rod to the complete heminested PCR mixture. The thermocycling steps were programmed to 85°C 1 cycle for 10 min, followed by 30 cycles of 85°C 30 sec for denaturing, 50°C 30 sec for annealing, 65°C 1 min for primer extension, and final extension 65°C for 10 minutes.

The crude heminested PCR products showing an amplicon of 398 bp in size at agarose gel electrophoresis were subject to Sanger reaction without further purification.

4.6. Primary and heminested PCR conditions for BRCA1 gene amplification

The primary PCR mixture contained 20 μL of ready-to-use LoTemp® PCR mix with denaturing chemicals (HiFi DNA Tech, LLC, Trumbull, CT, USA), 2 μL of water, 1 μL of 10 μmolar forward AG1 primer, 1 μL of 10 μmolar reverse R2 primer and 1 μL of sample RNA extract. For thermocycling, the temperature steps were programmed for an initial heating at 85 °C for 10 min, followed by 30 cycles at 85 °C for 30 s, 50 °C for 30 s and 65 °C for 1 min. The final extension was 65 °C for 10 min.

The heminested PCR mixture contained 20 μL of ready-to-use LoTemp® PCR mix, 1 μL of forward AG2 primer, 0.5 μL of reverse R2 primer and 3.5 μL of water in a total 25 μL volume. About 0.2 μL of the primary PCR products was transferred into the corresponding heminested PCR mixture with a micro-glass rod. The thermocycling steps were identical to those used for the primary PCR.

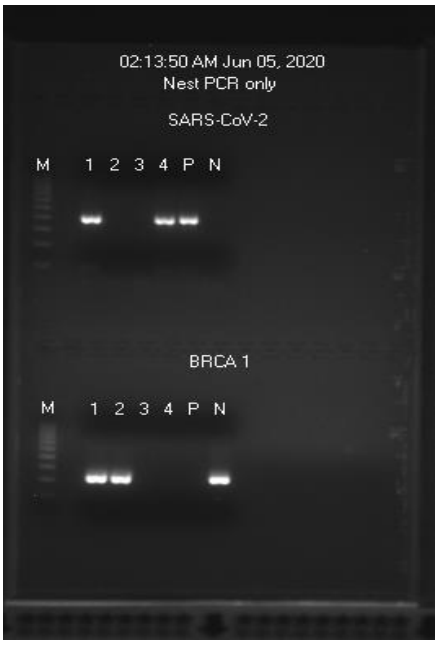
4.7. Interpretation of PCR results

As described above, 1 μL of RNA extract of patient sample, 1 μL of SARS-CoV-2 synthetic N gene RNA extract as positive control (P) and 1 μL of negative cell extract control (N) were used to initiate a set of primary PCR for viral nucleic acid detection.

In addition, 1 μL of each sample was used to initiate a *BRCA*1 gene primary PCR; and 1 μL of negative cell extract and 1 μL of water were used in the N and P control (see gel electrophoresis image below), respectively, for the *BRCA*1 primary PCR.

In routine testing, only the heminested PCR products were subjected to agarose gel electrophoresis because most primary PCR amplicons are not visible to the naked eyes.

An aliquot of 5 μL of each heminested PCR product was pipetted for agarose gel electrophoresis to detect the bands of the target DNA amplicons. As illustrated in the agarose gel image pasted below, for example, the negative control in Lane N of the SARS-CoV-2 series (upper half) and the water control in the P Lane of the *BRCA*1 (lower half) must show no amplicons. A 398-bp amplicon must be visualized in the P Lane of the SARS-CoV-2 series (upper half) and a 323-bp amplicon must be visualized in the N Lane of the *BRCA*1 series (lower half). Otherwise, the assay results are not valid.



As shown in the above image of gel electrophoresis, when the control amplicons on the P and N lanes are correctly visualized, there will be 4 possible combinations in the sample lanes. Sample 1 and 4 patterns are considered presumptive-positive for SARS-CoV-2 RNA. Sample 2 pattern is negative for SARS-CoV-2 as human *BRCA*1 gene is demonstrated in the sample. Sample 3 pattern indicates that no human cell genetic material was present in the sample being tested; the test result is invalid. Sample 4 pattern indicates that there is SARS-CoV-2 RNA in the test sample, but the human DNA has been extracted by the acidified phenol chloroform mixture. In some samples, there are more copies of viral genome equivalents than human genome copies. Since human genomic DNA is a potent PCR inhibitor, acidified phenol chloroform is used to remove most of the DNA from the sample to be tested.

The presumptive-positive crude heminested PCR products showing an amplicon of 398 bp in size in the SARS-CoV-2 series and the negative samples showing an amplicon of 323 bp in size in the *BRCA*1 series at agarose gel electrophoresis are subjected to Sanger reaction and automated sequencing for validation without further purification.

4.8. DNA Sequencing

The 398-bp SARS-CoV-2 heminested PCR product (about 0.2 μL), if detected at gel electrophoresis, was transferred by a micro-glass rod into a Sanger reaction tube containing 1 μL of 10 μmolar sequencing primer (Co3 or Co4), 1 μL of BigDye® Terminator (v 1.1/Sequencing Standard Kit), 3.5 μL 5× buffer, and 14.5 μL water in a total volume of 20 μL for 20 enzymatic primer extension/termination reaction cycles according to the protocol supplied by the manufacturer (Applied Biosystems, Foster City, CA, USA). For the SARS-CoV-2 negative samples, the 323-bp heminested PCR product was transferred by a micro-glass rod into a Sanger reaction tube containing 1 μL of 10 μmolar forward AG2 primer , 1 μL of BigDye® Terminator (v 1.1/Sequencing Standard Kit), 3.5 μL 5× buffer, and 14.5 μL water in a total volume of 20 μL for 20 enzymatic primer extension/termination reaction cycles according to the protocol supplied by the manufacturer (Applied Biosystems, Foster City, CA, USA). After a dye-terminator cleanup with a Centri-Sep column (Princeton Separations, Adelphia, NJ, USA), the reaction mixture was loaded in an Applied Biosystems SeqStudio Genetic Analyzer for sequence analysis. Sequence alignments were performed against the standard sequences stored in the GenBank database by on-line BLAST alignment analysis.

5. Conclusions

Testing the cellular components of respiratory tract specimens instead of cell-free fluids only and using conventional nested RT-PCR to amplify the target nucleic acid can reduce the number of false negatives in molecular testing for SARS-CoV-2. Routine sequencing of the nested PCR products not only can eliminate false positives, but also can categorize the positive isolate into one of the six clades of SARS-CoV-2 known to exist in the United States, based on single nucleotide polymorphisms in a 398-base segment of the N gene. Extremely accurate routine laboratory tests for SARS-CoV-2 are needed in long-term care facilities and as businesses attempt to return to normal operation. Molecular diagnosis of SARS-CoV-2 infection will become a serious matter and a positive test result must be substantiated by a DNA sequencing electropherogram showing the genomic fingerprints of the virus. False-positive test reports can easily create unnecessary panic resulting in negative impacts on local economies.

Based on an official document dated March 15, 2020, the FDA issued a letter on February 4, 2020 authorizing emergency use of the CDC 2019-Novel Coronavirus (2019-nCoV, *renamed as SARS-CoV-2*) Real-Time Reverse Transcriptase (RT)-PCR Diagnostic Panel for the presumptive qualitative detection of nucleic acid from the 2019-nCoV in upper and lower respiratory specimens [41]. The methodology presented in this paper is for definitive qualitative detection of nucleic acid from the SARS-CoV-2 infecting the cells in upper and lower respiratory specimens. The FDA also advises that in performing Clinical Evaluation of newly developed RT-qPCR assays for the SARS-CoV-2, using Recommended Comparator Method for percent agreement performance calculations, false results can be investigated using an additional EUA RT-PCR assay, and/or Sanger sequencing [42]. Routine nucleotide sequencing of all detected 398-bp N gene amplicons guarantees no-false positive results in SARS-CoV-2 assays.

Author Contributions

S.H.L.: Conceptualization, study design, methodology, data analysis, manuscript preparation and final review

Funding: None.

Acknowledgments: The author thanks the staff of the Connecticut State Department of Public Health, Microbiology Laboratory Division for the supply of 20 reference samples in support of developing a nested RT-PCR followed by

DNA sequencing test for SARS-CoV-2. The author also thanks Wilda Garayua for her technical assistance in performing the nucleotide sequencing.

Conflicts of Interest: Sin Hang Lee is Director of Milford Molecular Diagnostics Laboratory specialized in developing DNA sequencing-based diagnostic tests implementable in community hospital laboratories. The reagents and equipment used in this work can be substituted with similar products from various suppliers. The technology presented is not patentable.

References

1. Kelly, Meg, Sarah Cahlan and Elyse Samuels. 2020. What went wrong with coronavirus testing in the U.S. Washington Post. March 30, 2020. https://www.washingtonpost.com/politics/2020/03/30/11-100000-what-went-wrong-with-coronavirus-testing-us/
2. CDC. Public Health Response to the Initiation and Spread of Pandemic COVID-19 in the United States, February 24–April 21, 2020. Morbidity and Mortality Weekly Report (MMWR). Weekly / May 8, 2020 / 69;551-6. https://www.cdc.gov/mmwr/volumes/69/wr/mm6918e2.htm
3. Willman, David. 2020. CDC coronavirus test kits were likely contaminated, federal review confirms. Washington Post. June 20, 2020. https://www.washingtonpost.com/investigations/cdc-coronavirus-test-kits-were-likely-contaminated-federal-review-confirms/2020/06/20/1ceb4e16-b2ef-11ea-8f56-63f38c990077_story.html
4. Won J, Lee S, Park M, Kim TY, Park MG, Choi BY, Kim D, Chang H, Kim VN, Lee CJ. Development of a Laboratory-safe and Low-cost Detection Protocol for SARS-CoV-2 of the Coronavirus Disease 2019 (COVID-19). Exp Neurobiol. 2020; 29:107-119.
5. FDA. In Vitro Diagnostics EUAs. July 7, 2020. https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/vitro-diagnostics-euas
6. Rahman H, Carter I, Basile K, Donovan L, Kumar S, Tran T, Ko D, Alderson S, Sivaruban T, Eden JS, Rockett R, O'Sullivan MV, Sintchenko V, Chen SC, Maddocks S, Dwyer DE, Kok J. Interpret with caution: An evaluation of the commercial AusDiagnostics versus in-house developed assays for the detection of SARS-CoV-2 virus. J Clin Virol. 2020; 127:104374.
7. Park M, Won J, Choi BY, Lee CJ. Optimization of primer sets and detection protocols for SARS-CoV-2 of coronavirus disease 2019 (COVID-19) using PCR and real-time PCR. Exp Mol Med. 2020; 52: 963-977.
8. Elnifro EM, Ashshi AM, Cooper RJ, Klapper PE. Multiplex PCR: optimization and application in diagnostic virology. Clin Microbiol Rev. 2000; 13:559-570.
9. Katz AP, Civantos FJ, Sargi Z, Leibowitz JM, Nicolli EA, Weed D, Moskovitz AE, Civantos AM, Andrews DM, Martinez O, Thomas GR. False-positive reverse transcriptase polymerase chain reaction screening for SARS-CoV-2 in the setting of urgent head and neck surgery and otolaryngologic emergencies during the pandemic: Clinical implications. Head Neck. 2020; 42:1621-1628.
10. CDC. Interim Guidelines for Collecting, Handling, and Testing Clinical Specimens for COVID-19. https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.html
11. CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel. Instructions for Use. https://www.fda.gov/media/134922/download
12. QIAamp® Viral RNA Mini Handbook. January 2020.

13. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020; 579(7798):270-273.
14. Hockett RD, Kilby JM, Derdeyn CA, Saag MS, Sillers M, Squires K, Chiz S, Nowak MA, Shaw GM, Bucy RP. Constant mean viral copy number per infected cell in tissues regardless of high, low, or undetectable plasma HIV RNA. J Exp Med. 1999; 189:1545-54.
15. Adcock R, Cuzick J, Hunt WC, McDonald RM, Wheeler CM; New Mexico HPV Pap Registry Steering Committee. Role of HPV Genotype, Multiple Infections, and Viral Load on the Risk of High-Grade Cervical Neoplasia. Cancer Epidemiol Biomarkers Prev. 2019; 28:1816-1824.
16. FDA. Establishing the Performance Characteristics of In Vitro Diagnostic Devices for the Detection or Detection and Differentiation of Human Papillomaviruses. https://www.fda.gov/media/92930/download
17. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. Expert Rev Mol Diagn. 2020; 20:453-454.
18. Radomski N, Kreitmann L, McIntosh F, Behr MA. The critical role of DNA extraction for detection of mycobacteria in tissues. PLoS One. 2013; 8:e78749.
19. Bastien P, Procop GW, Reischl U. Quantitative real-time PCR is not more sensitive than "conventional" PCR. J Clin Microbiol. 2008; 46:1897-1900.
20. Goldenberger D, Künzli A, Vogt P, Zbinden R, Altwegg M. Molecular diagnosis of bacterial endocarditis by broad-range PCR amplification and direct sequencing. J Clin Microbiol. 1997; 35:2733-2739.
21. 新型冠状病毒肺炎实验室检测技术指南 http://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/jszl_11815/202003/W020200309540843062947.pdf
22. Lam TT, Jia N, Zhang YW, Shum MH, Jiang JF, Zhu HC, Tong YG, Shi YX, Ni XB, Liao YS, Li WJ, Jiang BG, Wei W, Yuan TT, Zheng K, Cui XM, Li J, Pei GQ, Qiang X, Cheung WY, Li LF, Sun FF, Qin S, Huang JC, Leung GM, Holmes EC, Hu YL, Guan Y, Cao WC. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*. 2020; 583(7815):282-285.
23. Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. Biomol Detect Quantif. 2015; 3:9-16.
24. Lalam N. Estimation of the reaction efficiency in polymerase chain reaction. J Theor Biol. 2006; 242:947-953.
25. Kramvis A, Bukofzer S, Kew MC. Comparison of hepatitis B virus DNA extractions from serum by the QIAamp blood kit, GeneReleaser, and the phenol-chloroform method. J Clin Microbiol. 1996; 34:2731-2733.
26. FDA. Cobas® HPV Test. https://www.accessdata.fda.gov/cdrh_docs/pdf10/p100020s017c.pdf
27. Lee SH. Guidelines for the use of molecular tests for the detection and genotyping of human papilloma virus from clinical specimens. Methods Mol. Biol. 2012; 903:65-101.
28. Lee SH, Vigliotti JS, Vigliotti VS, Jones W, Shearer DM. Detection of borreliae in archived sera from patients with clinically suspect Lyme disease. Int J Mol Sci. 2014; 15:4284-4298.
29. Pas SD, Patel P, Reusken C, Domingo C, Corman VM, Drosten C, Dijkman R, Thiel V, Nowotny N, Koopmans MP, Niedrig M. First international external quality assessment of molecular diagnostics for Mers-CoV. J Clin Virol. 2015; 69:81-85.

30. eCDC. Ebola virus disease ad hoc case definition for reporting in the EU. https://www.ecdc.europa.eu/en/all-topics-z/ebola-and-marburg-fevers/threats-and-outbreaks/ebola-outbreak-west-africa-2013-2016
31. Patel R, Babady E, Theel ES, Storch GA, Pinsky BA, St George K, Smith TC, Bertuzzi S. Report from the American Society for Microbiology COVID-19 International Summit, 23 March 2020: Value of Diagnostic Testing for SARS-CoV-2/COVID-19. mBio. 2020; 11:e00722-20.
32. CDC. Research Use Only 2019-Novel Coronavirus (2019-nCoV) Real-time RT-PCR Primer and Probe Information. https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html
33. Dyavar SR, Ye Z, Byrareddy SN, Scarsi KK, Winchester LC, Weinhold JA, Fletcher CV, Podany AT. Normalization of cell associated antiretroviral drug concentrations with a novel RPP30 droplet digital PCR assay. Sci Rep. 2018; 8:3626.
34. Imaizumi T, Yamamoto-Shimojima K, Yamamoto H, Yamamoto T. Establishment of a simple and rapid method to detect MECP2 duplications using digital polymerase chain reaction. Congenit Anom (Kyoto). 2020; 60:10-14.
35. FDA. Real-Time Fluorescent RT-PCR Kit for Detecting SARS-2019-nCoV https://www.fda.gov/media/136472/download
36. Peacock S. Health care: Bring microbial sequencing to hospitals. Nature 2014; 509(7502):557-9.
37. Peacock SJ, Weinstock GM. Microbial sequencing to improve individual and population health. Genome Med. 2014; 6:103.
38. Nao N, Shirato K, Katano H, Matsuyama S, Takeda M. Detection of second case of 2019-nCoV infection in Japan. 2020. https://www.niid.go.jp/niid/en/2019-ncov-e/9334-ncov-vir3-2.html
39. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CCS, Boshier FAT, Ortiz AT, Balloux F. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol. 2020; 83:104351.
40. Puranam RS, Attardi G. The RNase P associated with HeLa cell mitochondria contains an essential RNA component identical in sequence to that of the nuclear RNase P. Mol Cell Biol. 2001; 21:548-561.
41. Open letter from FDA to Robert R. Redfield, MD, Director, Centers for Disease Control and Prevention. March 15, 2020. https://www.fda.gov/media/134919/download
42. FDA. Molecular Diagnostic Template for Laboratories. Policy for Coronavirus Disease-2019 Tests During the Public Health Emergency (Revised) https://www.fda.gov/media/135659/download