Developing a Predictive Diagnostic for COVID-19

By

Christine Lu

Sc.B., Brown University, 2021

Thesis

Submitted in partial fulfillment of the requirements for the Degree of Master of Science in the Department of Molecular Pharmacology, Physiology, and Biotechnology at Brown University

PROVIDENCE, RHODE ISLAND

MAY 2023

AUTHORIZATION TO LEND AND REPRODUCE THE THESIS

As the sole author of this thesis, I authorize Brown University to lend it to other institutions for the purpose of scholarly research.

Date_____ _____
Christine Lu, Author

I further authorize Brown University to reproduce this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Date_____ _____
Christine Lu, Author

This thesis by Christine Lu is accepted in its present form by the Department of Molecular Pharmacology, Physiology, and Biotechnology as satisfying the thesis requirements for the degree of Master of Science

Date_____ Signature: _____

Dr. Gerard Nau, Advisor

Date_____ Signature: _____

Dr. Sean Monaghan, Reader

Date_____ Signature: _____

Dr. Alger Fredericks, Reader

Approved by the Graduate Council

Date_____ Signature: _____

Dr. Thomas A. Lewis, Dean of the Graduate School

Acknowledgements

I am extremely grateful to Dr. Gerard Nau for his mentorship throughout the program and for providing an environment that allowed me to grow intellectually and professionally.

I would also like to thank Dr. Sean Monaghan and Dr. Alger Fredericks for taking the time to join my committee and for generously supplying materials for my project including RNA samples and PAXgene tubes. In addition, I would like to thank Runping Zhao in Dr. Sean Monaghan's lab for providing invaluable help with isolating the RNA from patient blood samples and with RT-qPCR experiments.

Finally, I thank Allie Jacober, Talon Johnson, and Tanvi Patel in Dr. Gerard Nau's lab for all of their guidance and support, from showing me the ropes to helping out with experiments.

Table of Contents

Abstract of Developing a Predictive Diagnostic for COVID-19, by Christine Lu, ScM, Brown University, May 2023.

Objective: SARS-CoV-2 is an ongoing threat to human health. A predictive diagnostic would be helpful for providing information regarding the future progression of COVID-19 in hospitalized patients, however, none have been deployed. There are polymerase chain reaction (PCR) primers based on the sequences used by the CDC (CDC-N1) for diagnostics from nasopharyngeal swabs, but these primers are not effective for measuring RNAemia in all hospitalized COVID-19 patients. The objective of this project was to use deep RNA sequencing to develop a test with greater sensitivity than that of current diagnostics and provide a lead for developing a predictive diagnostic.

Methods: Primers targeting the ORF1ab and N genes were designed and validated through reverse transcription polymerase chain reaction (RT-PCR). The performance of our primers was compared to that of the CDC-N1. RNA was extracted from whole blood of COVID-19 patients admitted to the ICU. Primer designs were evaluated by reverse transcription quantitative polymerase chain reaction (RT-qPCR).

Results: An optimized RT-qPCR protocol was developed that minimizes background signal and detects SARS-CoV-2 sequences in whole blood. Higher levels of SARS-CoV-2 RNA were found in COVID-19 patients using our primers than with the CDC-N1 primers.

Conclusions: We found that our primers are consistently more sensitive than the CDC-N1 primers for the samples tested. Therefore, further studies of our primers should be conducted to elucidate their potential as a predictive diagnostic.

Introduction and Background

COVID-19 Pandemic

The COVID-19 pandemic is an ongoing global pandemic of the coronavirus disease 2019 (COVID-19). The first outbreak of COVID-19 occurred in the city of Wuhan, China in December 2019. From there, the coronavirus spread around the world through airborne particles and droplets, leaving devasting effects in its wake. The World Health Organization (WHO) officially declared the COVID-19 outbreak a global pandemic on March 11, 2020.[1] As of April 2023, there have been more than 700 million confirmed cases of COVID-19 and over six million deaths worldwide.[2] Furthermore, the COVID-19 pandemic has caused severe social and economic consequences globally, such as a large global recession that rivals the initial declines of the Great Depression.[3] There were widespread food and supply shortages due to both supply chain disruption and large-scale panic buying from consumers. In response to the pandemic, near-global lockdowns were initiated by countless countries and territories. Educational institutions transitioned to online learning, and many events were canceled or postponed, significantly disrupting typical pre-pandemic lifestyles.[4] As the COVID-19 pandemic continues to negatively impact human health and well-being, it is crucial to understand the best ways to treat COVID-19 in patients from all backgrounds in order to transition out of the global emergency as quickly as possible.

COVID-19 Disease

Patients with COVID-19 are infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and experience a spectrum of clinical manifestations as symptoms of the COVID-19 disease range from undetectable to critical. In particular, the National Institute of

Health (NIH) provides criteria for grouping patients with COVID-19 into the following categories according to the severity of illness: asymptomatic, mild illness, moderate illness, severe illness, and critical illness.[5] Severe and critical illness is most common in elderly or immunocompromised patients such as those with chronic lung disease, heart disease, cancer, Down Syndrome, and more.[6] COVID-19 transmission occurs when individuals inhale air that is contaminated by small airborne particles or droplets containing the coronavirus. The risk of transmission is higher when individuals are in close proximity to one another. However, SARS-CoV-2 can also transmit over longer distances, such as when individuals are indoors or in environments with low airflow.[7] Coronavirus transmission can also occur through contaminated fluids coming into contact with the eyes, nose, or mouth. COVID-19 patients that have mild to moderate illness typically remain contagious for up to 10 days, though this duration may be prolonged for those with more severe illness, and it is important to note that infected individuals are capable of spreading the virus even if no symptoms develop in the transmitter.[8] Lastly, mutations in the coronavirus have produced many variants, including Delta and Omicron, whose infectivity have further prolonged the COVID-19 pandemic.[9]

COVID-19 Prevention and Treatment

In response to the global pandemic, COVID-19 vaccines were approved and widely distributed in many countries. The preventive measures recommended by governing bodies include social distancing of at least six feet, wearing masks consistently when in the presence of others, and quarantining individuals who have been exposed or are symptomatic. Moreover, governments imposed travel restrictions, business restrictions and closures, quarantines, and contract tracing of infected individuals in an effort to ameliorate the

early stages of the COVID-19 pandemic.[10] While COVID-19 patients with mild illness can manage symptoms with over-the-counter medicines, patients who a have higher risk of progression to severe illness can be prescribed antiviral medications. These include the antiviral drug Veklury (Remdesivir) which was approved by the Food and Drug Administration (FDA) for hospitalized and non-hospitalized adult and pediatric patients.[11] Additionally, two oral antiviral pills, Paxlovid (Nirmatrelvir with Ritonavir) and Lagevrio (Molnupiravir), have been authorized for emergency use by the FDA for the treatment of adult patients with mild to moderate illness, though Paxlovid can also be prescribed to children age 12 and up.[11] Finally, there are two immune modulators**,** Olumiant (Baricitinib) and Actemra (Tocilizumab), that are FDA-approved for hospitalized adult patients.[11]

COVID-19 Diagnosis

The primary method of diagnosing SARS-CoV-2 in patients is through nucleic acid tests, of which the gold standard is currently considered to be reverse transcription polymerase chain reaction (RT-PCR).[12] These tests are most commonly performed using respiratory samples obtained through a nasopharyngeal swab, though a throat swab or sputum sample are other options as well.[13] Standard RT-PCR tests identify the presence of SARS-CoV-2 RNA by first converting the RNA into complementary DNA (cDNA) through the process called reverse transcription (RT). In the polymerase chain reaction (PCR), the cDNA, which serves as the template, is combined with PCR primers, which are used to identify and amplify the cDNA. Specifically, PCR primers are short, single-stranded segments of DNA that are used as a starting point for DNA synthesis during the amplification steps. A PCR primer set consists of a forward primer and a reverse primer that are designed to be complementary to the start and end of the

selected target sequence. During the PCR process, multiple cycles of amplification are carried out to produce a detectable amount of cDNA that can be qualitatively or quantitatively analyzed.[14]

Quantitative RT-PCR (RT-qPCR) experiments run on a real-time PCR instrument will produce cycle quantification (Cq) values, also known as the cycle threshold (Ct). These values represent the number of cycles necessary to detect the presence of PCR products above the background signal. An RT-qPCR test for COVID-19 is considered negative if no viral sequences are detected after 40 cycles of amplification.[15] The Cq values correlate inversely with the amount of starting material such that a low Cq value would indicate that there is a high level of the target sequence present in a sample. COVID-19 patients have the highest levels of SARS-CoV-2 during the first days of infection, so when the body of infected individuals works to clear the virus, the Cq values of subsequent tests would then be larger. However, though the assumption that there is a direct relation between raw Cq values and viral load is popular in recent studies, raw Cq values are not an accurate estimate of viral load. Therefore, data normalization by using a housekeeping gene as a reference is an important step during the interpretation of RT-qPCR results.[16]

While RT-qPCR tests are commonly used to determine the presence of SARS-CoV-2 for diagnosis, recent studies show that there is potential for RT-qPCR tests to provide prognostic information as well. For example, the levels of SARS-CoV-2 RNAemia, which is detectable SARS-CoV-2 RNA in blood, have been found to be correlated with disease severity and patient mortality when measured in plasma or serum samples using RT-qPCR.[17,18] The detection of SARS-CoV-2 RNA in nasopharyngeal samples using RT-qPCR was similarly shown to be associated with disease severity and patient mortality.[19] Furthermore, the levels of SARS-CoV-2

RNA in nasopharyngeal samples have a dose-dependent correlation with RNAemia and mortality risk as demonstrated through RT-qPCR.[20] While multiple studies have shown that RT-qPCR tests can correlate with disease severity and mortality, there have not been any deployed tests that offer prognostic value. A limitation of the tests that have been studied is that they do not consistently detect RNAemia in all COVID-19 patient samples. Therefore, a test with enhanced sensitivity that is capable of reliably detecting SARS-CoV-2 in all COVID-19 patients would provide more information on prognosis as more sensitivity would allow for detecting a spectrum of severity that would help with categorizing patients. As such, there exists a need for developing a predictive diagnostic for COVID-19 that measures SARS-CoV-2 RNAemia with greater sensitivity and would be more helpful at providing prognostic information than current tests available.

SARS-CoV-2 Genome

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the pathogen that causes COVID-19. It is a single-stranded and positive-sense RNA virus that belongs to the *Coronaviridae* family and shares 96% genome similarity with a bat coronavirus.[21] SARS-CoV-2 contains a genome of around 30 kilobases, which is one of the largest genomes among all RNA virus families. The SARS-CoV-2 virus has fourteen open reading frames (ORFs), that encode a total of twenty-seven different proteins. Its genome includes a 5′ untranslated region (UTR), replication complex (ORF1ab), Spike (S) gene, Envelope (E) gene, Membrane (M) gene, Nucleocapsid (N) gene, 3′ UTR, several unidentified non-structural ORFs, and a poly (A) tail.[22] The SARS-CoV-2 virus replicates rapidly in the cells of infected patients which initiates multiple signaling cascades, including activation of the pro-inflammatory and antiviral responses that

trigger a “cytokine storm.” The virus then rapidly disseminates through peripheral blood to other organs like the heart, kidney, liver, or spleen.[23]

The ORF1ab and N genes are good targets in studies of the SARS-CoV-2 genome as they are highly conserved, and their sequences are unique for SARS-CoV-2.[24] The value of the ORF1ab and N genes as diagnostic targets is supported by the deep RNA sequencing of patients with COVID-19. In a single-center, prospective study of fifteen patients with COVID-19 that were admitted to the intensive care unit (ICU), SARS-CoV-2 RNA was identified in the blood of all of the patients using deep RNA sequencing.[25] The majority of the reads that aligned to the SARS-CoV-2 genome corresponded to the non-structural protein (NSP) RNA dependent RNA polymerase and N genes.[25] The most common sequences found in the reads will be referred to as the NSP12-Peak and N-Peak respectively.[25] The RNA dependent RNA polymerase gene is located in ORF1ab of the SARS-Cov-2 genome and encodes an enzyme that catalyzes the replication of RNA from an RNA template.[26] The N gene encodes the N protein, which is an RNA-binding protein that is critical for viral replication, genome packaging and helps to increase the efficiency of virus transcription and assembly.[27]

During the COVID-19 pandemic, the Centers for Disease Control and Prevention (CDC) provided sequences of PCR primers for diagnostic testing that specifically target the N gene of SARS-CoV-2.[28] These primers will be referred to as CDC-N1, and it is important to note that the region to which the CDC-N1 primers bind is not the same as the N-Peak (Figure 1). However, the CDC-N1 primers have not been shown to be capable of detecting SARS-CoV-2 in the blood of all hospitalized COVID-19 patients.[29] Therefore, we hypothesize that a diagnostic test using primers targeting the NSP12-Peak or N-Peak sequences identified through deep RNA

sequencing will provide better sensitivity than diagnostic tests targeting other areas in the SARS-CoV-2 genome.

The research described here is the optimization of a real-time RT-qPCR protocol implementing primers that are designed to target the N-Peak region of SARS-CoV-2. In addition, we evaluated the performance of our N-Peak primers as compared to the CDC-N1 primers. The long-term goal is to develop this protocol into a predictive diagnostic for COVID-19 that is capable of identifying patients who are most likely to progress to severe illness.

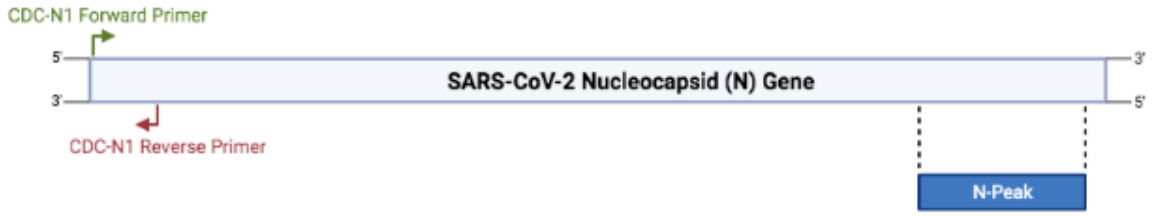

Figure 1: SARS-CoV-2 Nucleocapsid Gene Structure (Created with BioRender.com)

Materials and Methods

COVID-19 Patient Cohort

Patients with COVID-19 who were admitted to the Intensive Care Unit (ICU) at Rhode Island Hospital were retrospectively enrolled. Participants had documented evidence of SARS-CoV-2 infection based on a positive PCR test using samples obtained from the nasopharynx.[25] The population and characteristics of ICU patients are as previously described, and blood samples were collected on day 0 of ICU admission.[25]

Primers for SARS-CoV-2 and Human Genes

The sequences of the primers used in this study are listed in Table 1. The NSP1, NSP12-Peak, N-ORF, and N-Peak primers were purchased from Integrated DNA Technologies (USA). The CDC-N1 primer sequences were obtained from the Centers for Disease Control and Prevention (CDC) and purchased from Sigma Aldrich (USA). The beta-actin (ACTB) primers were purchased from Invitrogen (USA), and the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) primers were purchased from Integrated DNA Technologies (USA).

Table 1: SARS-CoV-2 and Human Primer Sequences

Name	**Target Gene**	**Forward Primer Sequence 5'-3'**	**Reverse Primer Sequence 5'-3'**
NSP1	SARS-CoV-2 ORF1ab gene	TACGGCGCCGATCTAAAGTC	CACGGGTAACACCACTGCTA
NSP12-Peak	SARS-CoV-2 ORF1ab gene	ACCCTCACCTTATGGGTTGG	ACACGTTGTATGTTTGCGAGC
N-ORF	SARS-CoV-2 Nucleocapsid gene	ACCCGCAATCCTGCTAACAA	ACGAGAAGAGGCTTGACTGC
N-Peak	SARS-CoV-2 Nucleocapsid gene	ATGAAACTCAAGCCTTACCGCA	ATCCAAATCTGCAGCAGGAAG
CDC-N1	SARS-CoV-2 Nucleocapsid gene	GACCCCAAAATCAGCGAAAT	TCTGGTTACTGCCAGTTGAATCTG
ACTB	Human beta-actin gene	GCACCACACCTTCTACAATGAG	ATAGCACAGCCTGGATAGCAAC
GAPDH	Human GAPDH gene	TCAAGGCTGAGAACGGGAAG	CGCCCCACTTGATTTTGGAG

End-Point RT-PCR and Gel Electrophoresis

Reverse transcription (RT) of SARS-CoV-2 genomic RNA (BEI Resources, USA) was performed using the SuperScript IV First-Strand Synthesis System (Invitrogen, USA) following

the manufacturer's instructions. One hundred nanograms of SARS-CoV-2 genomic RNA was reverse transcribed to complementary DNA (cDNA) in a final volume of 20µL, and the cDNA was stored at -20°C until use. Then, polymerase chain reaction (PCR) was performed using DreamTaq DNA Polymerase (Thermo Scientific, USA) following the manufacturer's instructions. Each 25µL PCR reaction included primers at a final concentration of 400nM and 12.5ng of cDNA. The PCR protocol was run on a Mastercycler Personal thermal cycler (Eppendorf, USA) with the following conditions: 95°C for 2 minutes, 35 cycles of steps: 1) 95°C for 30 seconds; 2) 52°C for 30 seconds; 3) 72°C for 1 minute, and 72°C for 5 minutes. PCR products were visualized in 2% agarose gels stained with SYBR Safe.

RNA Isolation from PAXgene Tubes

Blood samples from patients with COVID-19 who presented to the emergency room (ER) were obtained from the Brown/Lifespan COVID-19 Biorepository. The blood samples were collected in PAXgene Blood RNA Tubes (PreAnalytiX, USA), which contain a proprietary reagent that allows for immediate stabilization of intracellular RNA in whole blood, and were stored at -80°C until use. For RNA isolation, these samples were processed with the PAXgene Blood RNA Kit IVD (PreAnalytiX, USA) following the manufacturer's protocol.

Total blood RNA from patients with COVID-19 who were admitted to the intensive care unit (ICU) was extracted from blood samples stored in PAXgene Blood RNA Tubes as described previously and was provided by Dr. Sean Monaghan.[25]

The concentration of all RNA samples was determined using the QuantiFluor RNA System (Promega, USA) according to the manufacturer's instructions.

Real-Time RT-qPCR

The RNA isolated from COVID-19 patient blood samples was reverse transcribed to cDNA using the SuperScript IV First-Strand Synthesis System (Invitrogen, USA) following the manufacturer's instructions in a final volume of 20µL, and the cDNA was stored at -20°C until use. Real-time qPCR was performed using iTaq Universal SYBR Green Supermix (Bio-Rad, USA) following the manufacturer's instructions. The final volume of each qPCR reaction was 10µL, and the final concentration of the primers in each reaction was 400nM. All qPCR reactions were centrifuged at 455 RCF for 1 minute prior to thermal cycling. The qPCR was performed using a CFX Connect or CFX96 instrument (Bio-Rad, USA) controlled by the CFX Maestro Software (Bio-Rad, USA) with the following thermal cycling protocol: 95°C for 30 seconds and 40-50 cycles of steps: 1) 95°C for 5 seconds; 2) 60°C for 30 seconds. The cycle quantification (Cq) values generated by the CFX system were analyzed using the $2^{-\Delta\Delta Cq}$ method of relative quantification. With this method, delta Cq values were first calculated to normalize the Cq values for the gene of interest to the ACTB housekeeping gene (Table 2). Next, delta delta Cq values were calculated using the delta Cq value for reactions with CDC-N1 primers as the calibrator to which reactions with N-Peak primers are compared (Table 2). Lastly, the $2^{-\Delta\Delta Cq}$ equation was used to calculate relative fold difference values that were then log-transformed (Table 2).

Table 2: Values in Relative Quantification Analysis

Value	**Calculation**
Delta Cq (ΔCq)	ΔCq = Cq (N-Peak or CDC-N1) – Cq (ACTB)
Delta Delta Cq (ΔΔCq)	ΔΔCq = ΔCq (N-Peak) – ΔCq (CDC-N1)
Relative Fold Difference	relative fold difference = $2^{-(\Delta\Delta Cq)}$
Log_2 Relative Fold Difference	$\log_2$ relative fold difference = $\log_2(2^{-(\Delta\Delta Cq)})$

Results

Confirmation of PCR Primers

In the first phase of this work, PCR primers targeting the ORF1ab and Nucleocapsid (N) genes of SARS-CoV-2 were designed. Specifically, the NSP1 and NSP12-Peak primers target the ORF1ab gene while the N-ORF and N-Peak primers target the N gene. To verify that the primers worked as predicted, end-point RT-PCR and gel electrophoresis were performed using the methods described. The RNA template used in RT-PCR was purified by BEI Resources from the cell lysate and supernatant of *Cercopithecus aethiops* kidney epithelial cells that were infected with SARS-CoV-2 isolate USA-WA1/2020. PCR products produced by the NSP1, NSP12-Peak, N-ORF, and N-Peak primers are 109, 104, 114, and 104 base pairs long respectively. We found that the PCR products produced by each primer set were the expected sizes when visualized through gel electrophoresis and estimated using the reference ladder (Figure 2). Therefore, we concluded that the NSP1, NSP12-Peak, N-ORF, and N-Peak primers effectively amplify the correct PCR products.

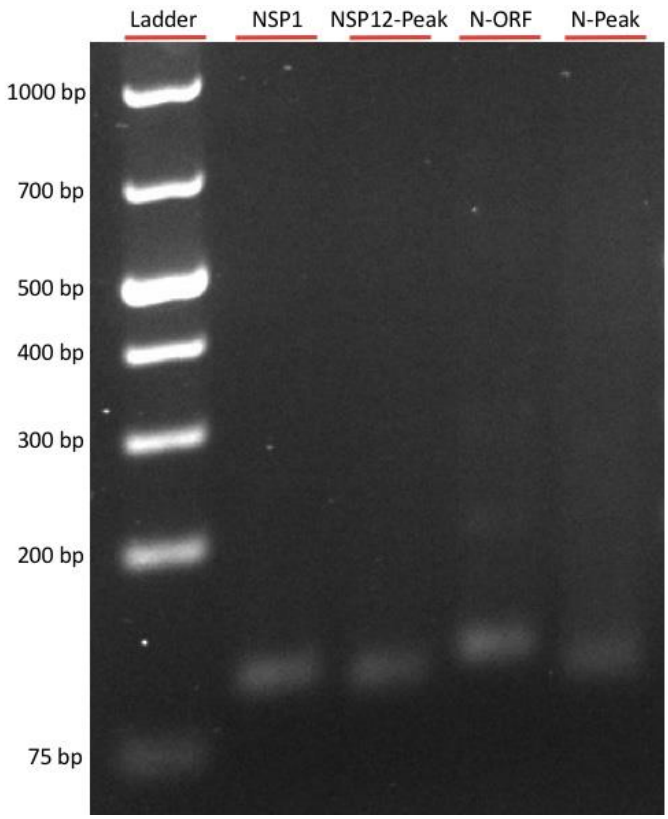

Figure 2: Visualization of PCR Products Produced by Primers Targeting SARS-CoV-2. PCR products in the NSP1, NSP12-Peak, N-ORF, and N-Peak lanes are of larger size than the primers themselves or primer-dimers. High molecular weight products of uncertain origin were present in the N-ORF lane and are nonsignificant to subsequent experiments.

Preliminary RT-qPCR Experiments

To compare the sensitivity of the N-Peak and CDC-N1 primers, real-time RT-qPCR using RNA from COVID-19 patients admitted to the ER or ICU was performed as described in the methods. For the preliminary experiments, the positive control was the GAPDH housekeeping gene. In addition, two negative controls were included in each experiment. A No-Template-Control (NTC) was used to monitor reagent contamination while a No-Reverse-Transcriptase (NRT) control was used to detect genomic DNA contamination in the RNA.

Multiple experiments performed on different days and with different samples produced consistent results that demonstrated a need for optimizing the RT-qPCR protocol. There were three main issues that occurred throughout the preliminary experiments. The first issue was that PCR products were consistently detected between 25 and 30 cycles for the NRT control with GAPDH primers (Figure 3). The qPCR protocol for the preliminary experiments included a total of 50 cycles in which no PCR products were detected for the NTCs with GAPDH primers. We expect no PCR products to be detected in the NRT control as well as the NTC, so the detection of PCR products in less than 30 cycles for the NRT control suggests that there is genomic DNA contamination present. The second issue was that PCR products were consistently detected between 35 and 40 cycles for the NTCs with N-Peak primers (Figure 3). This indicates that there was contamination that occurred during the preparation and set-up of the qPCR. Lastly, the third issue was that PCR products were consistently detected between 35 and 40 cycles for NRT controls with the N-Peak primers while there were no PCR products detected in the NRT controls with the CDC-N1 primers (Figure 3). This demonstrates that there was a high background signal in the NRT controls with N-Peak primers specifically. Therefore, in order to produce results that can be properly analyzed, these three issues first had to be resolved.

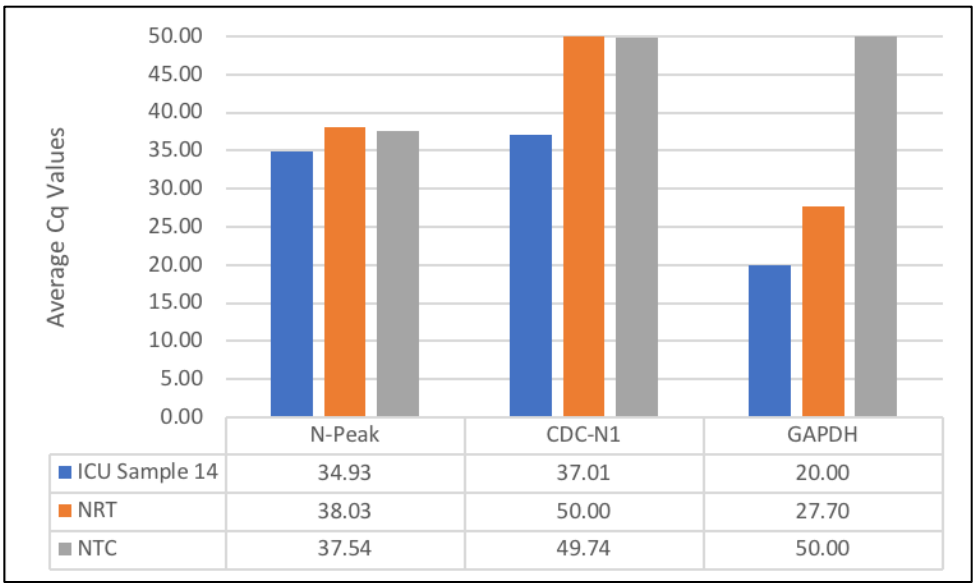

Figure 3: Average Cycle Quantification (Cq) Values for ICU Sample 14 in Preliminary Experiments. Cq values were determined by the CFX Maestro software for ICU Sample 14 (blue), No-Reverse-Transcriptase Controls (orange), and No-Template-Controls (gray). Cq values of 50.00 means no PCR products were detected.

Optimization of RT-qPCR Protocol: Positive Control

In order to reduce the contamination in the NRT control for GAPDH, all efforts were made to eliminate environmental contamination during the production of cDNA. This included performing protocols within a biosafety cabinet that was cleaned using 10% bleach and UV radiation. We also obtained new reagents, microfuge tubes, and filter tips that were only opened inside the biosafety cabinet. Pipettes were decontaminated with 10% bleach and were likewise only used within the biosafety cabinet. Aliquots of all new reagents were made to reduce the possibility of contaminating stock reagents. However, despite these precautions, the Cq values for NRT controls with GAPDH primers in subsequent experiments remained between 25 and 30 cycles (Figure 4). For comparison, we obtained primers for beta-actin (ACTB), another housekeeping gene, and found that PCR products were not detected until 40 to 50 cycles for the NRT control with ACTB primers (Figure 4). Given that the ACTB primers seem to produce a

lower background signal than the GAPDH primers, ACTB was used as the positive control for all subsequent experiments.

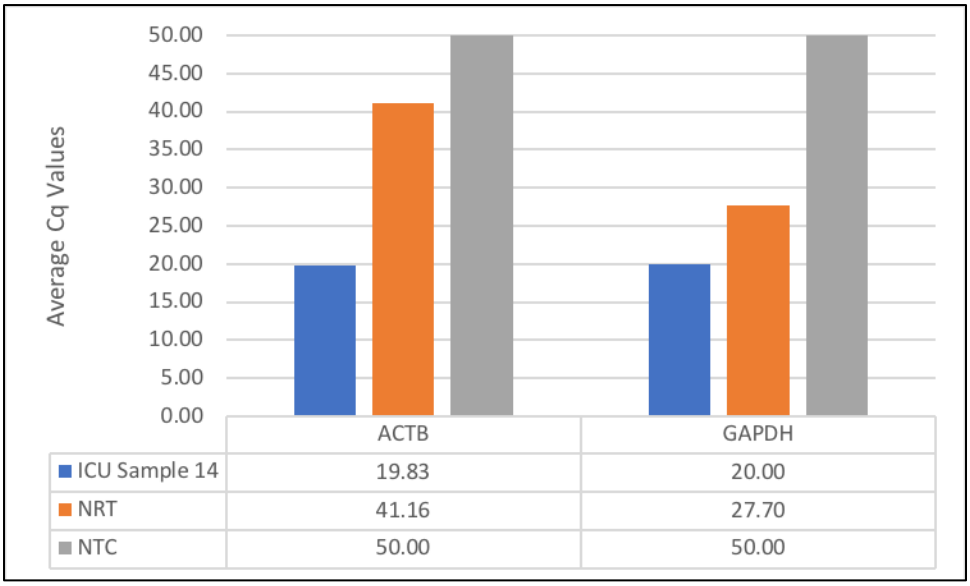

Figure 4: Average Cycle Quantification (Cq) Values of Housekeeping Genes ACTB and GAPDH for ICU Sample 14. Cq values were determined by the CFX Maestro software for ICU Sample 14 (blue), No-Reverse-Transcriptase Controls (orange), and No-Template-Controls (gray). Cq values of 50.00 means no PCR products were detected.

Optimization of RT-qPCR Protocol: No-Template-Control

To address the contamination of the NTCs with N-Peak primers in the following experiments, we maintained use of the biosafety cabinet with the cleaning and operational parameters previously described. However, the Cq values for the NTCs remained similar to that of the preliminary experiments, which indicated that the source of contamination may not be from the environment. In order to test whether internal contamination among wells may be present, we re-designed the plate map. In the preliminary experiments, the plate was setup such that the experimental sample, NRT control, and NTC wells for the N-Peak primers were all

placed in one row. In the re-designed plate map, a row was dedicated to only NRT controls while a different row was dedicated to only NTCs such that the negative control wells were not placed in the same row as the experimental sample wells. When performing subsequent experiments utilizing the new plate map, we found that there were no PCR products detected in the NTCs with N-peak primers, which indicates that the detection of PCR products in the NTCs from preliminary experiments was likely due to cross-well contamination (Figure 5). While the contamination in the NTCs with N-Peak primers improved significantly with the implementation of the re-designed plate map, there was little to no improvement in the NRT controls with N-Peak primers, which suggests that there were other factors contributing to the high background signal (Figure 5).

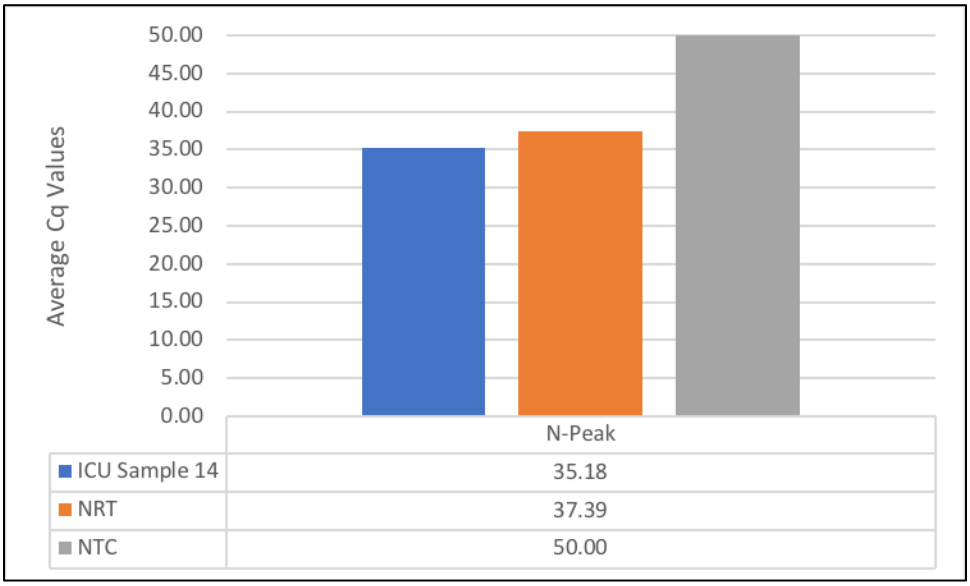

Figure 5: Average Cycle Quantification (Cq) Values for ICU Sample 14 in Experiments with the Re-Designed Plate Map. Cq values were determined by the CFX Maestro software for ICU Sample 14 (blue), No-Reverse-Transcriptase Controls (orange), and No-Template-Controls (gray). Cq values of 50.00 means no PCR products were detected.

Optimization of RT-qPCR Protocol: No-Reverse-Transcriptase Control

To lower the background signal of the NRT controls with the N-Peak primers, we investigated the design of our N-Peak primers. We hypothesized that obtaining new N-Peak primers which mimic the CDC-N1 primers in design may help to reduce the background signal of the NRT controls with the N-Peak primers given that the CDC-N1 primers did not produce high background signal in the NRT controls of preliminary experiments. We noticed that the biggest difference in primer design between the N-Peak and CDC-N1 primers was the GC content, which is the percentage of nitrogenous bases that are guanine (G) or cytosine (C) in a primer sequence. The CDC-N1 forward and reverse primers had 45% and 46% GC content respectively while the N-Peak forward and reverse primers had 55% and 52% GC content respectively. Therefore, we re-designed the N-Peak primers so that the forward and reverse primers have 45% and 48% GC content respectively, and the N-Peak sequences listed in Table 1 reflect this re-design. Simultaneously we shortened the qPCR protocol from 50 cycles of amplification to 40 cycles since 40 cycles is considered to be the industry standard. After these interventions, we found that there were no PCR products detected for the NRT controls with N-Peak primers (Figure 6). As the results of the shorter, 40-cycle experiments demonstrate minimized background signal with the re-designed N-Peak primers, we determined this version to be the optimized protocol, which is utilized for all experiments performed thereafter. To summarize, the optimized protocol includes the following specifications: 1) replacing GAPDH with ACTB as the positive control; 2) separating the NRT controls and NTCs into different rows; 3) replacing the high-background N-Peak primers with the re-designed N-Peak primers shown in Table 1; 4) changing the qPCR cycling parameters from 50 cycles to 40 cycles. With the

implementation of the optimized protocol, we found that the three main issues identified during the preliminary experiments have been resolved.

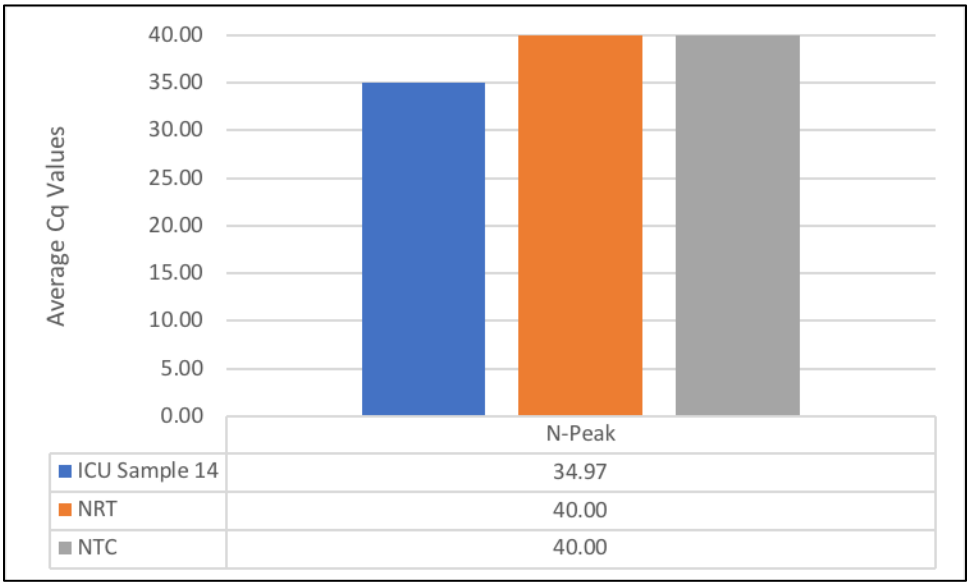

Figure 6: Average Cycle Quantification (Cq) Values for ICU Sample 14 in Experiments with the Optimized Protocol. Cq values were determined by the CFX Maestro software for ICU Sample 14 (blue), No-Reverse-Transcriptase Controls (orange), and No-Template-Controls (gray). Cq values of 40.00 means no PCR products were detected.

N-Peak vs. CDC-N1 Performance for ICU Patients with COVID-19

To evaluate the N-Peak primer design in comparison with the CDC-N1 primers for patients with COVID-19 that were admitted to the ICU, real-time RT-qPCR experiments were performed using the optimized protocol. In all RT-qPCR experiments using RNA from ICU patients with COVID-19, the N-Peak PCR products were detected between 32 and 39 cycles. For example, with ICU Sample 1, the N-Peak products were detected at an average of 35.15 cycles while no amplicons were detected using CDC-N1 (Figure 7). Therefore, the N-Peak primers were more sensitive than the CDC-N1 primers as the N-Peak primers were able to detect the

presence of SARS-CoV-2 when the CDC-N1 primers did not. This pattern was also observed for ICU Samples 6, 7, 8, 12, 13, and 15. In other ICU samples that were tested with the optimized RT-qPCR protocol, amplicons from the CDC-N1 primers were detected between 34 and 40 cycles. However, even when CDC-N1 products were detected, the corresponding Cq value was consistently higher than that of the N-Peak, which indicates that the N-Peak primers were more sensitive than the CDC-N1 primers since there were more N-Peak products generated than CDC-N1 products during the qPCR protocol (Figure 8). For all samples that were tested, the ACTB PCR products were consistently detected around 16 to 20 cycles.

To analyze N-Peak primer performance relative to CDC-N1 performance, the log-transformed relative fold difference values were calculated as described in the methods. This value represents how much more, or less, sensitive the N-Peak primers were compared to the CDC-N1 primers. Experiments in which excessive contamination was detected using the NRT controls and NTCs were excluded from the analysis. In general, the N-Peak primers outperformed the CDC-N1 primers for all samples that were tested. The $\log_2$ relative fold difference values range from 1.13 to 6.84 for the samples shown, which indicates that the N-Peak primers are approximately 2-fold to 100-fold more sensitive than the CDC-N1 primers in these samples (Figure 9).

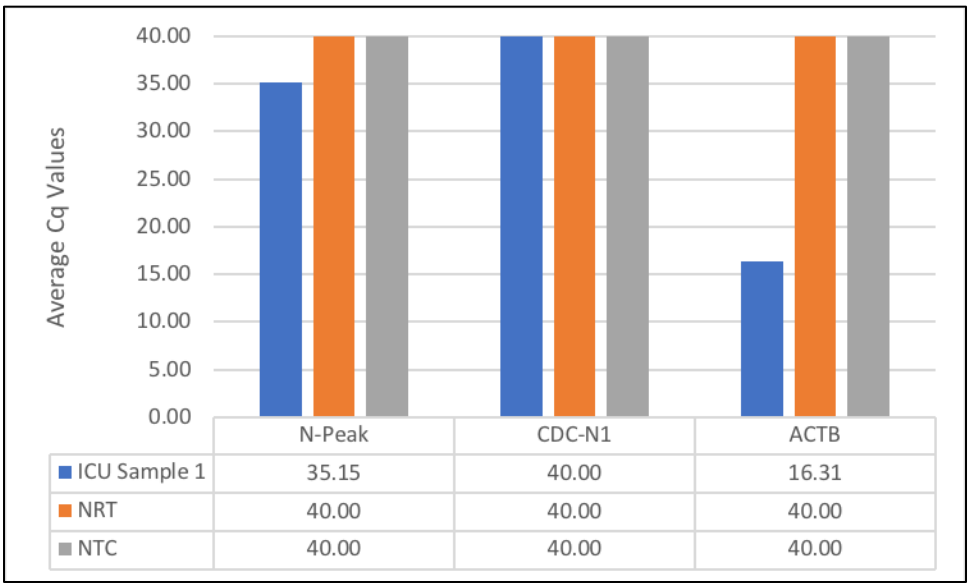

Figure 7: Average Cycle Quantification (Cq) Values for ICU Sample 1. Cq values were determined by the CFX Maestro software for ICU Sample 1 (blue), No-Reverse-Transcriptase Controls (orange), and No-Template-Controls (gray). Cq values of 40.00 means no PCR products were detected.

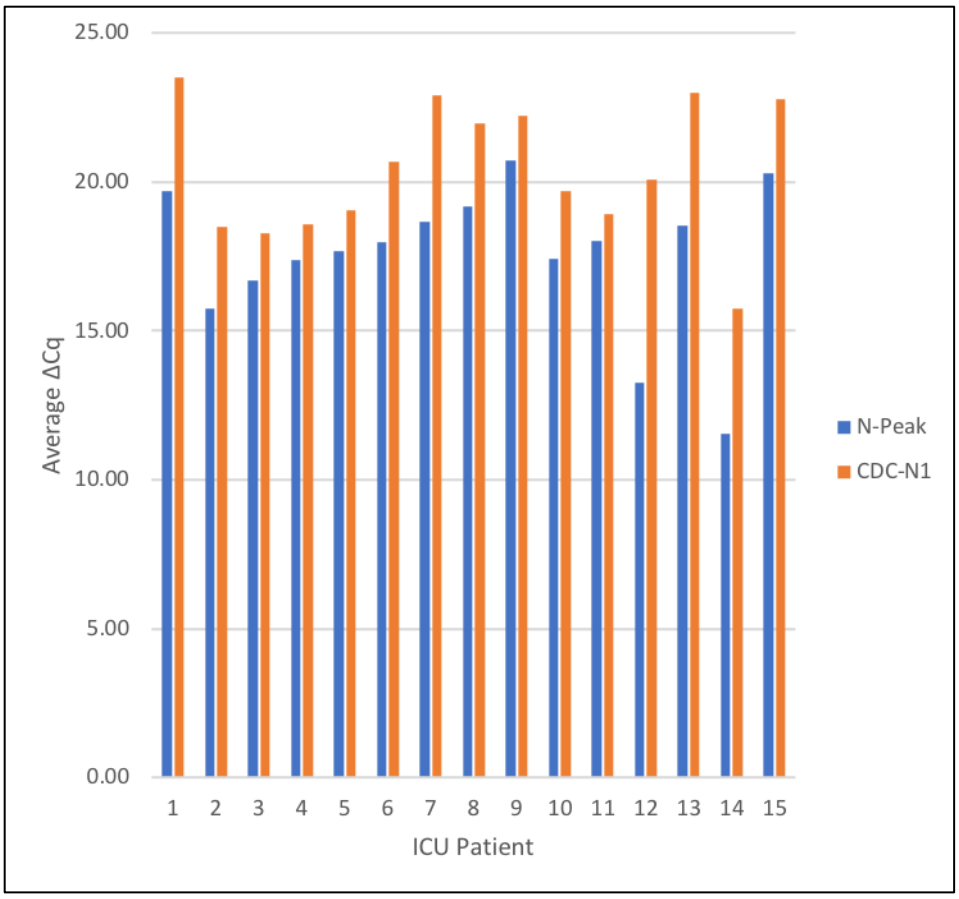

Figure 8: Average Delta Cq (ΔCq) Values for ICU Patients with COVID-19. Delta Cq values were calculated as described in the Methods section. Delta Cq values represent normalization of N-Peak (blue) and CDC-N1 (orange) Cq values to ACTB Cq values.

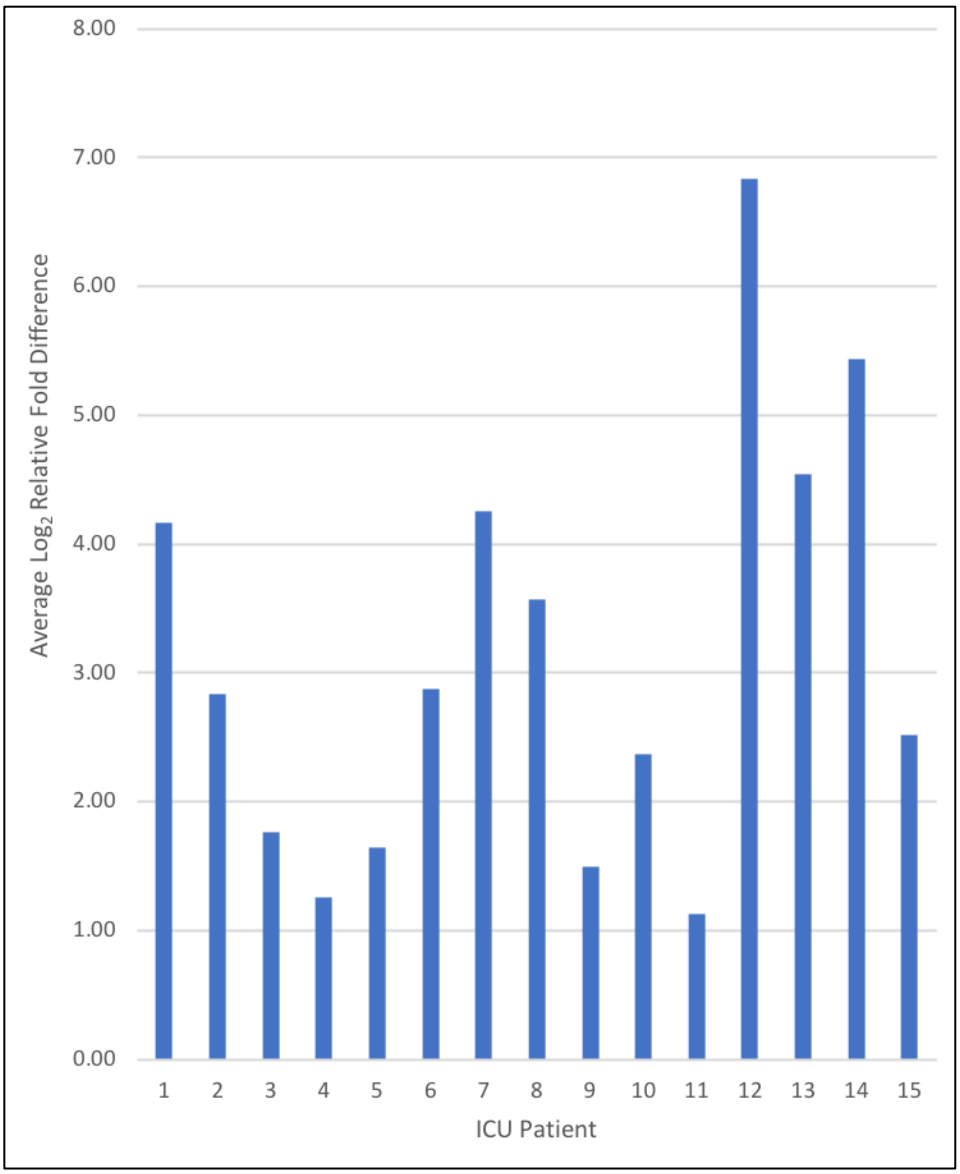

Figure 9: Average Log_2 Relative Fold Difference Values for ICU Patients with COVID-19. Fold difference between detection of SARS-CoV-2 using N-Peak versus CDC-N1 primers were calculated as described in the Methods section. A log_2 relative fold difference greater than one means the N-Peak primers are more sensitive than the CDC-N1 primers.

Discussion and Conclusions

SARS-CoV-2 remains an ongoing threat to human health as numerous variants continue to arise and spread throughout communities. A predictive diagnostic test would be helpful for clinicians that manage hospitalized COVID-19 patients, but though some tests exist that correlate

with disease severity, none have been deployed that offers prognostic value. This project investigates a lead for the development of a test with enhanced sensitivity that would provide more information on prognosis. Deep RNA sequencing has identified the NSP12-Peak and N-Peak sequences within the ORF1ab and N genes that are most commonly detected in the blood of COVID-19 patients admitted to the ICU.[25] Our work makes use of primers targeting the N-Peak sequence to provide an optimized RT-qPCR protocol that detects SARS-CoV-2 in whole blood and generates data with minimal background signal for analysis. The final, optimized protocol can be utilized for future RT-qPCR experiments using RNA isolated from blood stored in PAXgene tubes or in various other applications.

Furthermore, this study supports our hypothesis that using primers targeting the N-Peak will provide better sensitivity than the CDC-N1 primers in a diagnostic test using blood samples. Therefore, a test that implements the N-Peak primers could provide a more accurate measurement of RNAemia that could be helpful for clinicians in determining how to treat a patient for COVID-19. Our work demonstrates that deep RNA sequencing can be used to inform PCR primer design that results in a more sensitive assay. However, whether an N-Peak diagnostic test is capable of providing prognostic information remains to be determined.

A future direction for this project could be to perform the optimized RT-qPCR protocol using samples obtained from the Brown/Lifespan COVID-19 Biorepository of patients who were confirmed to have SARS-CoV-2 infection and presented to the Emergency Room (ER) at Rhode Island Hospital. Comparing the relative fold difference values between ER and ICU patients could help determine whether the severity of COVID-19 in patients can be assessed using a test with the N-Peak primers. Additionally, it would be beneficial to correlate the relative fold difference values for ICU patients with the number of N-Peak reads detected in each patient

through deep RNA sequencing. This would allow us to determine if the extent of N-Peak sensitivity could be directly related to the levels of N-Peak reads in the blood. Further, the clinical characteristics of the ICU patients whose samples were tested in this study can be investigated to determine if there are any commonalities that may explain the difference in N-Peak sensitivity between samples. Lastly, the optimized RT-qPCR protocol can be performed with the NSP12-Peak primers for the COVID-19 patients admitted to the ICU in order to investigate how the sensitivity of those primers compares to that of the N-Peak primers. In conclusion, we have found the N-Peak sequence to be a promising target for the measurement of RNAemia, and consequent studies of the N-Peak primers should be conducted to elucidate its potential as a predictive diagnostic.

References

1. *Coronavirus disease (COVID-19) pandemic*. (n.d.). Retrieved April 1, 2023, from https://www.who.int/europe/emergencies/situations/covid-19
2. *WHO Coronavirus (COVID-19) Dashboard*. (n.d.). Retrieved April 1, 2023, from https://covid19.who.int
3. *How Does the Pandemic Recession Stack Up against the Great Depression?* (n.d.). Retrieved April 1, 2023, from https://www.stlouisfed.org/on-the-economy/2020/october/pandemic-recession-stack-great-depression
4. Onyeaka, H., Anumudu, C. K., Al-Sharify, Z. T., Egele-Godswill, E., & Mbaegbu, P. (2021). COVID-19 pandemic: A review of the global lockdown and its far-reaching effects. *Science Progress*, *104*(2), 00368504211019854. https://doi.org/10.1177/00368504211019854
5. *Clinical Spectrum*. (n.d.). COVID-19 Treatment Guidelines. Retrieved April 1, 2023, from https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/
6. *COVID-19: Who's at higher risk of serious symptoms?* (n.d.). Mayo Clinic. Retrieved April 1, 2023, from https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-who-is-at-risk/art-20483301
7. *Transmission of SARS-CoV-2: Implications for infection prevention precautions*. (n.d.). Retrieved April 1, 2023, from https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions
8. Walsh, K. A., Spillane, S., Comber, L., Cardwell, K., Harrington, P., Connell, J., Teljeur, C., Broderick, N., de Gascun, C. F., Smith, S. M., Ryan, M., & O'Neill, M. (2020). The

duration of infectiousness of individuals infected with SARS-CoV-2. *The Journal of Infection*, *81*(6), 847–856. https://doi.org/10.1016/j.jinf.2020.10.009

9. Vasireddy, D., Vanaparthy, R., Mohan, G., Malayala, S. V., & Atluri, P. (2021). Review of COVID-19 Variants and COVID-19 Vaccine Efficacy: What the Clinician Should Know? *Journal of Clinical Medicine Research*, *13*(6), 317–325. https://doi.org/10.14740/jocmr4518
10. CDC. (2023, January 26). *COVID-19 and Your Health*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html
11. Research, C. for D. E. and. (2023, March 10). *Coronavirus (COVID-19) | Drugs*. FDA; FDA. https://www.fda.gov/drugs/emergency-preparedness-drugs/coronavirus-covid-19-drugs
12. *SARS-CoV-2 Testing*. (n.d.). COVID-19 Treatment Guidelines. Retrieved April 1, 2023, from https://www.covid19treatmentguidelines.nih.gov/overview/sars-cov-2-testing/
13. CDC. (2020, February 11). *Labs*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.html
14. Bustin, S. A., & Nolan, T. (2020). RT-qPCR Testing of SARS-CoV-2: A Primer. *International Journal of Molecular Sciences*, *21*(8), 3004. https://doi.org/10.3390/ijms21083004
15. Sethuraman, N., Jeremiah, S. S., & Ryo, A. (2020). Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA*, *323*(22), 2249–2251. https://doi.org/10.1001/jama.2020.8259

16. Miranda, R. L., Guterres, A., de Azeredo Lima, C. H., Filho, P. N., & Gadelha, M. R. (2021). Misinterpretation of viral load in COVID-19 clinical outcomes. *Virus Research*, *296*, 198340. https://doi.org/10.1016/j.virusres.2021.198340

17. Jacobs, J. L., Bain, W., Naqvi, A., Staines, B., Castanha, P. M. S., Yang, H., Boltz, V. F., Barratt-Boyes, S., Marques, E. T. A., Mitchell, S. L., Methé, B., Olonisakin, T. F., Haidar, G., Burke, T. W., Petzold, E., Denny, T., Woods, C. W., McVerry, B. J., Lee, J. S., … Mellors, J. W. (2022). Severe Acute Respiratory Syndrome Coronavirus 2 Viremia Is Associated With Coronavirus Disease 2019 Severity and Predicts Clinical Outcomes. *Clinical Infectious Diseases*, *74*(9), 1525–1533. https://doi.org/10.1093/cid/ciab686

18. Rodríguez-Serrano, D. A., Roy-Vallejo, E., Zurita Cruz, N. D., Martín Ramírez, A., Rodríguez-García, S. C., Arevalillo-Fernández, N., Galván-Román, J. M., Fontán García-Rodrigo, L., Vega-Piris, L., Chicot Llano, M., Arribas Méndez, D., González de Marcos, B., Hernando Santos, J., Sánchez Azofra, A., Ávalos Pérez-Urria, E., Rodriguez-Cortes, P., Esparcia, L., Marcos-Jimenez, A., Sánchez-Alonso, S., … REINMUN-COVID Group. (2021). Detection of SARS-CoV-2 RNA in serum is associated with increased mortality risk in hospitalized COVID-19 patients. *Scientific Reports*, *11*(1), 13134. https://doi.org/10.1038/s41598-021-92497-1

19. Kawasuji, H., Morinaga, Y., Tani, H., Yoshida, Y., Takegoshi, Y., Kaneda, M., Murai, Y., Kimoto, K., Ueno, A., Miyajima, Y., Fukui, Y., Kimura, M., Yamada, H., Sakamaki, I., & Yamamoto, Y. (2022). SARS-CoV-2 RNAemia with a higher nasopharyngeal viral load is strongly associated with disease severity and mortality in patients with COVID-19. *Journal of Medical Virology*, *94*(1), 147–153. https://doi.org/10.1002/jmv.27282

20. Hagman, K., Hedenstierna, M., Widaeus, J., Arvidsson, E., Hammas, B., Grillner, L., Jakobsson, J., Gille-Johnson, P., & Ursing, J. (2022). Correlation of SARS-CoV-2 Nasopharyngeal CT Values With Viremia and Mortality in Adults Hospitalized With COVID-19. *Open Forum Infectious Diseases*, *9*(9), ofac463. https://doi.org/10.1093/ofid/ofac463

21. Rastogi, M., Pandey, N., Shukla, A., & Singh, S. K. (2020). SARS coronavirus 2: From genome to infectome. *Respiratory Research*, *21*(1), 318. https://doi.org/10.1186/s12931-020-01581-z

22. Shah, A., Rashid, F., Aziz, A., Jan, A. U., & Suleman, M. (2020). Genetic characterization of structural and open reading Fram-8 proteins of SARS-CoV-2 isolates from different countries. *Gene Reports*, *21*, 100886. https://doi.org/10.1016/j.genrep.2020.100886

23. Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., Liu, S., Zhao, P., Liu, H., Zhu, L., Tai, Y., Bai, C., Gao, T., Song, J., Xia, P., Dong, J., Zhao, J., & Wang, F.-S. (2020). Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *The Lancet. Respiratory Medicine*, *8*(4), 420–422. https://doi.org/10.1016/S2213-2600(20)30076-X

24. Alanagreh, L., Alzoughool, F., & Atoum, M. (2020). The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. *Pathogens*, *9*(5), Article 5. https://doi.org/10.3390/pathogens9050331

25. Fredericks, A. M., Jentzsch, M. S., Cioffi, W. G., Cohen, M., Fairbrother, W. G., Gandhi, S. J., Harrington, E. O., Nau, G. J., Reichner, J. S., Ventetuolo, C. E., Levy, M. M., Ayala, A., & Monaghan, S. F. (2022). Deep RNA sequencing of intensive care unit

patients with COVID-19. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-20139-1

26. Romano, M., Ruggiero, A., Squeglia, F., Maga, G., & Berisio, R. (2020). A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. *Cells*, *9*(5), 1267. https://doi.org/10.3390/cells9051267
27. Bai, Z., Cao, Y., Liu, W., & Li, J. (2021). The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. *Viruses*, *13*(6), 1115. https://doi.org/10.3390/v13061115
28. *2019-novel coronavirus (2019-nCoV) real-time rRT-PCR panel primers and probes*. (n.d.). Retrieved April 1, 2023, from https://stacks.cdc.gov/view/cdc/84525
29. Fajnzylber, J., Regan, J., Coxen, K., Corry, H., Wong, C., Rosenthal, A., Worrall, D., Giguel, F., Piechocka-Trocha, A., Atyeo, C., Fischinger, S., Chan, A., Flaherty, K. T., Hall, K., Dougan, M., Ryan, E. T., Gillespie, E., Chishti, R., Li, Y., … Li, J. Z. (2020). SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nature Communications*, *11*, 5493. https://doi.org/10.1038/s41467-020-19057-5