Host, Viral, and Environmental Transcriptome Profiles of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

Authors:
Daniel J. Butler[1#], Christopher Mozsary[1#], Cem Meydan[1,3,4#], David Danko[1,2#], Jonathan Foox[1,3#], Joel Rosiene[5,6], Alon Shaiber[5,6], Ebrahim Afshinnekoo[1,3,4], Matthew MacKay[1], Fritz J. Sedlazeck[7], Nikolay A. Ivanov[1,3,8], Maria Sierra[1], Diana Pohle[13], Michael Zietz[14], Undina Gisladottir[14], Craig D. Westover[1], Krista Ryon[1], Benjamin Young[1], Chandrima Bhattacharya[1], Phyllis Ruggiero[6], Bradley W. Langhorst[15], Nathan Tanner[15], Justyna Gawrys[6], Dmitry Meleshko[1,2], Dong Xu[10], Jenny Xiang[8,10], Angelika Iftner[13], Daniela Bezdan[13], John Sipley[6], Lin Cong[6], Arryn Craney[6], Priya Velu[6], Ari M. Melnick[16], Iman Hajirasouliha[1,3,17], Thomas Iftner[13], Mirella Salvatore[9], Massimo Loda[6], Lars F. Westblade[6,9], Shawn Levy[11], Melissa Cushing[6], Nicholas Tatonetti[14], Marcin Imielinski[5,6*], Hanna Rennert[6*], Christopher E. Mason[1,3,5,12*]

Affiliations:
[1]Department of Physiology and Biophysics, Weill Cornell Medicine, NY, USA.
[2]Tri-Institutional Computational Biol. & Medicine Program, Weill Cornell Medicine, NY, USA
[3]The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medicine, NY, USA
[4]WorldQuant Initiative for Quantitative Prediction, Weill Cornell Medicine, NY, USA
[5]New York Genome Center, NY, USA
[6]Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, NY, USA
[7]Baylor College of Medicine, Houston, TX, USA
[8]Clinical & Translational Science Center, Weill Cornell Medicine, NY, USA
[9]Division of Infectious Diseases, Department of Medicine, Weill Cornell Medicine, NY, USA
[10]Genomics Resources Core Facility, Weill Cornell Medicine, NY, USA
[11]HudsonAlpha Discovery Institute, Huntsville, AL, USA
[12]The Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, NY, USA
[13]Institute of Medical Virology and Epidemiology of Viral Diseases, University Hospital Tuebingen, Germany
[14]Department of Biomedical Informatics, Department of Systems Biology, Department of Medicine, Institute for Genomic Medicine, Columbia University, NY, USA
[15]New England Biolabs, MA, USA
[16]Department of Medicine, Weill Cornell Medicine, NY, USA
[17]Englander Institute for Precision Medicine and the Meyer Cancer Center, Weill Cornell Medicine, NY, USA

[#]These authors contributed equally to this work
*To whom correspondence should be addressed

Corresponding Authors:
Marcin Imielinski (mai9037@med.cornell.edu); Hanna Rennert (har2006@med.cornell.edu); Christopher E. Mason (chm2042@med.cornell.edu)

Abstract
A global pandemic of the 2019 coronavirus disease (COVID-19), caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), is responsible for over 150,000 deaths worldwide, including >12,000 deaths in New York City (NYC) alone. Given the rapid emergence of this pathogen, little is known about its genetic variation, immune system interactions, population prevalence, and environmental distribution. As a result, there is a pressing clinical and public health need for scalable molecular technologies that can rapidly detect SARS-CoV-2 infection and robustly interrogate strain evolution and host response in patients. To address these challenges, we designed a loop-mediated isothermal amplification (LAMP) assay to identify SARS-CoV-2 infection within 30 minutes of application, including directly from lysed cells. Simultaneously, we developed a large-scale host and viral transcriptomic profiling platform that employs total RNA sequencing (RNA-seq) of nasopharyngeal swab specimens. Applying both technologies to 442 samples, spanning 338 clinical samples tested for SARS-CoV-2, 86 environmental samples from the NYC subway, and 14 controls, we assembled a broad molecular picture of the COVID-19 epidemic in NYC. We found close concordance between viral titers measured with our rapid LAMP assay, RNA-seq, and the state-of-the-art RT-PCR. Full transcriptomic analyses revealed a distinct subset of the Western European clade A2a as the predominant genomic subtype in NYC cases, which was defined by two single nucleotide variants in nsp2 and ORF 3a and a clade-specific, 9-bp in-frame deletion in the virulence factor nsp1. High SARS-CoV2 viral titers were associated with distinct host transcriptional responses, including activation of the *ACE2* gene and interferon response genes (e.g., *IFIT1*). Since ACE inhibitors (ACEIs) can also increase *ACE2* expression, outcomes for patients taking ACEIs were examined, and showed a significantly increased risk of intubation and death. Our results demonstrate the utility of two molecular diagnostic platforms in defining the genetic features of an evolving global pandemic and provide insights that can aid future COVID-19 diagnostics, public health monitoring, and therapeutic options.

Keywords: severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), 2019 coronavirus disease (COVID-19), loop-mediated isothermal amplification (LAMP), real-time polymerase chain reaction (RT-PCR), next-generation sequencing (NGS), RNA-seq, global health

Introduction

In March 2020, a novel viral pandemic was declared by the World Health Organization (WHO) for the 2019 coronavirus disease (COVID-19), an infection caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (He *et al*., 2020). New York City (NYC) emerged as a global hotspot of this pandemic, with over 120,000 cases, 32,000 hospitalizations, and 12,000 deaths attributed to COVID-19 as of April 17, 2020, according to John's Hopkins Coronavirus Resource Center (https://coronavirus.jhu.edu).

The presenting symptoms of COVID-19 resemble those of common viral respiratory infections, requiring molecular diagnosis to reliably distinguish SARS-Cov-2 infection from influenza and the common cold (Guan *et al.,* 2020, Zhou *et al*., 2020). Current approaches to molecular testing are mostly limited to hospital labs and are largely reserved for the most severe cases, with limited accessibility to the general population. As a result, the prevalence of SARS-Cov-2 in the population is mostly unknown, particularly among mild or asymptomatic cases. Though several novel, scalable biotechnological innovations for viral testing have begun to emerge (e.g.

CRISPR-Cas12a (Broughton *et al*., 2020) or CRISPR-Cas13 (Metsky *et al*., 2020) on paper-based detection systems, or loop-mediated isothermal amplification (LAMP, Tanner *et al*., 2015, Zhang *et al.*, 2020, Yu *et al*., 2020, Schmid-Burgk *et. al*., 2020) these have not been validated against gold-standard clinical assays or next-generation sequencing (NGS).

The lack of rapid and widely available SARS-Cov-2 diagnostics has fundamentally limited the public health approach to COVID-19, including the implementation of contact tracing and estimation of infection fatality rates. In addition, the persistence of SARS-CoV-2 across a range of surfaces (e.g. glass, metal) for up to nine days in experimental conditions (van Doremalen *et al*., 2020) and hospital areas (Ong *et al*, 2020) raises the possibility that fomite transmission may be playing a role in COVID-19 spread. A key question is whether the environmental surface distribution of SARS-Cov-2 in high-traffic areas (e.g. subways) may have driven its rapid emergence in certain regions (e.g. NYC).

Modern viral surveillance combines large-scale molecular diagnostics with networks of data sharing (Gardy *et al*., 2015). Resources such as Global Initiative on Sharing All Influenza Data (GISAID), the JHU Dashboard, and NextStrain, have enabled dynamic tracking of the evolution of the COVID-19 pandemic (Dong *et al*., 2020, Meyers *et al*., 2020, Hadfield *et al*., 2018). Nevertheless, the fraction of cases that have undergone full-length viral sequencing are still low (<0.5% of documented cases), underscoring the need for additional profiling, particularly in global infection epicenters such as NYC. Full-length viral sequences are required to chart the dynamics of strain evolution, infer the temporal and geographic trajectories of spread, and correlate clinical features (e.g. disease severity, comorbidities, viral load) with specific genotypes. Indeed, such genomic epidemiology efforts have shown trajectories of the spread of SARS-CoV-2 in Washington, California, and Connecticut (Zhao *et al*., 2020, Gonzalez-Reiche *et al*., 2020, Fauver *et al*., 2020) as well as in China (Lu *et al*., 2020). However current approaches to viral profiling (targeted sequencing of SARS-Cov-2) fail to provide information on host immune response or microbial co-infections, whose associations could exacerbate or modify presentation of COVID-19.

To address these gaps in technology and disease knowledge, a rapid LAMP assay was designed and optimized to detect SARS-Cov-2 infection, including from saliva cells' lysate. Then, an NGS platform was established to comprehensively profile nasopharyngeal swabs from patient samples with total RNA sequencing (RNA-seq). Both technologies were applied to 338 COVID-19 patients at Weill Cornell Medicine / New York Presbyterian (WCM/NYP) hospital and 86 environmental samples obtained from high-transit areas in the NYC subway in early March. These host and viral expression data were then compared to clinical metadata from NYP, specifically with a focus on ACE inhibitors and outcomes. Our results propose novel molecular approaches for detecting and tracking SARS-Cov-2 infection, provide new molecular insights into the evolution of the COVID-19 outbreak in NYC, and implicate specific host factors and drug interactions for the disease.

Results

SARS-CoV-2 LAMP Assay Validation

We developed a colorimetric assay to quickly detect and quantify SARS-CoV-2 viral load in patient and environmental samples, utilizing a set of six LAMP primers and simple single tube protocol (**Figure 1b**). Primers were designed to create two nested loops and amplify within the SARS-CoV-2 nucleocapsid gene (N gene), which enabled a 30-minute reaction workflow. Related pathogens, high prevalence disease agents, and normal or pathogenic flora that are reasonably likely to be encountered in the clinical specimens were evaluated to identify the % homology between the primer sequences and these other organisms, and the probes were also designed to avoid known polymorphisms (see Methods).

To validate the assay, we first evaluated two synthetic RNAs (see Methods) whose sequences matched the viral sequences of patients from Wuhan, China and Melbourne, Australia (**Supp. Fig. 1**). The first control (MT007544.1) was used to test the analytical sensitivity via the limit of detection (LoD), titrated from 1 million molecules of virus (10^6) down to a single copy, using serial log-10 dilutions. The reaction output was measured at 0-, 20-, and 30-minute intervals (**Figure 1c**) before the samples were heated to 95°C for inactivation. The LoD was found to be between 10–100 viral total copies, and this was then replicated to show a similar LoD of 10–100 viral copies on the second synthetic control from the patient from Wuhan (**Supp. Figure 2b**).

We evaluated the reproducibility, sensitivity, and specificity for the LAMP assay across a set of serial titrations. Across these experiments, LAMP fluorescence correlated closely with SARS-CoV-2 RNA viral copies (**Figure 1d**), with an overlap of the median signal from negative controls at lower levels (25 total copies) of viral RNA (n=10). This translated to a 100% reproducibility at 1,000 and 500 total copies of viral RNA, 90% reproducibility at 100 copies, and 80% reproducibility at 50 copies (**Supp. Figure 3a**). This indicates an LoD threshold (95% reproducibility at two times the LoD) that is likely near 200 copies of RNA, with a maximum sensitivity of 10-50 copies of the viral RNA.

To optimize the LAMP assay for clinical samples, we tested it on a range of reaction types, dilutions, and reaction volumes. As expected, clinically positive samples showed a much higher fluorescence than negative samples, with superior performance observed with higher (10µL vs 5µL) input volumes (**Supp. Figure 3b**). Clinically positive samples that failed to generate LAMP fluorescence were associated with lower viral genome abundance (high cycle threshold [Ct] value qRT-PCR). We obtained similar performance on bulk saliva lysate (**Supp. Figure 3c-e**), including increasing reaction sensitivity as a function of viral copy number. These results indicate robust performance of the LAMP assay across a broad range of cellular lysates (see Methods).

Having optimized the LAMP assay, we evaluated its efficacy as a diagnostic across 133 samples that tested positive (clinical positives) and 205 samples that tested negative (clinical negatives) for SARS-CoV-2 by qRT-PCR (see Methods). Among these, 182 had sufficient total nucleic acids for testing. Quantitative LAMP fluorescence data were used to build a Receiver Operator Characteristic (ROC) plot and evaluate the clinical sensitivity and specificity of the assay. At the optimal threshold, we observed an overall sensitivity of 94% and specificity of 100% (**Figure 1e**) as well as a positive predictive value (PPV) of 95% and negative predictive value (NPV) of 100%. With increasing viral load, as measured by Ct values, the LAMP assay showed an increased diagnostic sensitivity, with the lowest sensitivity (78%) up to the highest (100%),

especially for those Ct values showing the highest level of virus (15-25 Ct values, **Supp. Fig. 4**). These same LAMP assay thresholds yielded consistent test positivity for synthetic RNA positive controls (Twist Biosciences) as well as clinical spike-in carrier RNAs (20/20) 12/12) (**Supp. Fig. 4).**

SARS-CoV-2 Genomes and Phylogenetics

To further characterize these COVID-19 patients, shotgun metatranscriptomics (RNA-sequencing with ribosomal RNA depletion) was performed for both COVID-POS and COVID-NEG patient samples (mean= 63.2M read pairs), which sequences human, microbial, and viral RNAs. The proportion of reads that mapped to the SARS-CoV-2 genome was high and low for the positive and negative controls, respectively (**Figure 2a)**. Samples annotated as positive by RT-PCR had higher viral content (0.1%-62.2% of total RNA) than the RT-PCR-negative or environmental samples. A high overall correlation (R^2= 0.78) was observed between the NGS viral abundance and the quantification by RT-PCR and LAMP (**Figure 2b, Supp. Figure 5**). Indeed, the samples which tested negative by RT-PCR (n=205) were almost all below 0.01% of total reads, except for three NGS samples that indicated the presence of the SARS-CoV-2 virus by read proportion (>0.01%), high coverage (>99%) of the SARS-CoV-2 genome, and which also tested positive with LAMP (**Supp. Figure 6**); another three showed lower evidence of the SARS-CoV-2 genome (>97% coverage). These data provide evidence that false negatives from RT-PCR are likely rare (3-6/205, or 1.4-2.8%).

The NGS reads were then aligned to several NCBI databases and examined for other organisms using Kraken2 (Wood *et al*., 2019). All samples showed reads that mapped to either human, bacterial, or SARS-CoV-2 genomes, and then more rarely (in order) reads that mapped to other viruses, fungi, and archaea (**Figure 2c, 2d**). The reads that mapped to SARS-CoV-2 from total RNA-seq were sufficient to provide >10x coverage of the viral genomes for 118/133 (88%) of patient samples (**Figure 3a**), and also was relatively robust to sequencing depth for the high viral titer samples. Indeed, when downsampling the sets of reads/patient (50M down to 0.5M), we observed consistent detection of the viral genome across the patients' RNA (**Figure 3b**).

All samples with sufficient coverage across the 30kb SARS-CoV-2 genome (>10x) were then used to call genetic variants, and NYC viral strains were placed into the phylogenetic tree of the global outbreak. Single nucleotide variants (SNVs) and indels (insertions/deletions) were called with FreeBayes relative to the SARS-CoV-2 reference sequence (GCF_009858895.2) and then compared with 4,964 sequenced samples in GISAID. By using the maximum likelihood phylogenetic tree built from the NextStrain database (Hadfield *et al*., 2018), a global range of likely sources of the virus were found. However, the most common SNVs were found in alleles that were over-represented in the European cluster of cases, as well as in alleles that indicated Asian and American origin (**Figure 4a**); these results confirm those of other NYC strains deposited into GISAID (Gonzalez-Reiche *et al*., 2020).

The most common sites of viral genetic variation (>10% of samples) in the genome observed across the NYC strains matched the most variable sites from the global viral database in GISAID (**Figure 4a**). This includes many sites with 70-80% prevalence across our samples (e.g. 3037C>T, 23403 A>G), as well as matched proportions for the lower (~20%) allele frequencies (e.g. 8782C>T). The majority of NY cases are found to fall within clade A2a. Three single

nucleotide variants (1059C>T, 11916C>T, and 25563G>T) were found to be significantly (FDR < 0.1) enriched for NY cases within A2a as compared with non-NY cases (**Figure 4a**). Of the three SNVs, 25563G>T was found to delineate a putative subclade within clade A2a. 260 (~60%) of the NY cases belong to this subclade, but these NY cases only constitute 24.9% of the cases in this A2a1 and the rest are predominantly from Europe. The rest of the 163 (~40%) NY cases are spread over rest of the phylogenetic tree with no additional large clusters.

Previous work has indicated the possibility of sub-strains of the SARS-CoV-2 virus (Tang *et al*., 2020), based on mutations in the gene for open reading frame 8 (ORF8, site 84). These two strains (L and S) have previously shown some geospatial and temporal stratification, indicating possible drift and/or selection of the strains, although the functional or virulence significance is not clear (Tang *et al*., 2020). The potentially more virulent "L type" was reported to be more prevalent in the early stages of the outbreak in Wuhan (70% of cases), but then decreased after early January 2020 (Tang *et al*., 2020). An examination of all NYC strains from this work demonstrated an even stronger trend, with a greater prevalence (90%) of the L strain overall, which represent the early stages of the NYC outbreak (**Figure 4b,** March 10-16, when the caseload rose from 186 to 5,096 confirmed cases and deaths from 0 to 19 total).

Larger variants were also found across the viral genomes, including a 9-bp deletion found in two patients, in position 684 of the open reading frame (ORF) gene 1b (ORF1b, in non-structural protein 1, NSP1). This deletion was also observed as the most common large deletion in the GISAID database (**Figure 4c**), including samples from England, Iceland, and Canada. This deletion spans three amino acids and includes a tyrosine that is not present in the SARS-CoV-1 genome (**Figure 4d**), and thus represents a C-terminal deletion and variant of this NSP gene that is unique to CoV-2.

Since the total RNA-seq sequences mapped to other organisms, it was possible to characterize the presence of other infections or commensal species, including the subset of patients who tested negative by RT-PCR for SARS-CoV-2. A range of other RNA viruses and organisms were found that were distinct from the high, medium, and lower viral titer patients (**Supp Table 1**), with overall similarity seeing across patients in the bacterial RNA fractions (**Figure 5a)**. However, there were distinct variations in known respiratory viruses, including human coronaviruses 229E, NL63, and HKU1, as well as Influenza A, Rhinovirus (A and C), and respiratory syncytial virus (**Figure 5a)**. Most patients that presented with the SARS-CoV-2 virus were absent of other respiratory viruses, but a few patients (4/133, 3.0%) also carried Influenza A, and this was most often observed in the lower titer (bottom third) of COVID-POS (3/4) patients. One unfortunate patient carried SARS-CoV-2, metapneumovirus, and coronavirus NL63 at the same time (**Figure 5a)**.

Using these same analytical methods, 86 environmental samples that were collected on the NYC subway between March 6 to 13, 2020 were examined for SARS-CoV-2. Handrails, kiosks, and the floor of Grand Central and Times Square subway stations were swabbed using a sterile and DNA/RNA free swab, following the MetaSUB sampling protocols for nucleic acid stabilization (Danko *et al*., 2020). These samples were then tested with the LAMP assay and total RNA-seq. As expected, environmental samples from underground subways showed a greater number of reads mapping to fungal and archaeal species (Danko *et al*., 2020) (**Supp. Table 1**), but showed

no evidence of SARS-CoV-2 (**Figure 2)**. Indeed, none of the subway samples reached significant levels of virus based on the proportion of reads (**Figure 2a**) or total reads, which was especially clear with the dual-index library preparations (**Supp Figure 7**). However, a broad range of other bacterial and viral species was found, including a large set of phages (e.g. Streptomyces phage VWB), desiccation-tolerant bacteria (e.g *Deinococcus radiodurans*), and more abundant bacterial and archaeal RNA than the clinical samples (**Supp. Table 1**). Of note, these were checked against a database of putative false positives (**Supp. Table 2**), which was created from *in silico* fragmentation of the SARS-CoV-2 genome and mapping against the same database.

Defining the SARS-CoV-2 Host Transcriptome

The RNA-sequencing reads that mapped to the host genome were then used to quantify the host response to the SARS-CoV-2 infection in the epithelial cells collected from the NP swabs. Human-mapped reads were examined with DESeq2 to discern the differentially expressed genes (DEGs) between COVID-POS and COVID-NEG patients (see Methods). Overall, 5,982 significant DEGs (q<0.01, >1.5-fold change) were found for the COVID-POS samples (**Supp. Table 3**), which matched with the quantitative and clinical data from RT-PCR, NGS, and LAMP (**Figure 6a**). This spanned 2,942 up-regulated DEGs and 3,040 down-regulated DEGs, which could be separated into high, medium, and low viral response categories based on the three detection methodologies (RT-PCR, NGS, and LAMP) (**Figure 6b**); the host transcriptome that exhibited the greatest amount of DEGs were those with the highest viral titer.

Differentially expressed host genes indicated a wide range of antiviral responses, including some a common interferon response across all ranges of viral levels (**Figure 6b**). Notably, host epithelial cells showed an increase in angiotensin converting enzyme 2 (*ACE2)* expression (q-value=0.006), which is the SARS-CoV-2 cellular receptor. This critical gene for viral entry exhibited a dose-dependent expression concomitant with the higher levels of SARS-CoV-2 virus, along with IFI27 and IFI6 (Interferon Alpha Inducible Protein 6) (**Figure 6c**). Other interferon-related genes included IFIT1, an interferon-induced antiviral RNA-binding protein that specifically binds single-stranded RNA bearing a 5'-triphosphate group (PPP-RNA) and SHFL (Shiftless Antiviral Inhibitor of Ribosomal Frameshifting), which is associated with Dengue virus and Chikungunya infection. The DEGs also included HERC6 (HECT Containing E3 Ubiquitin Protein Ligase Family Member 6), which aids Class I MHC-mediated antigen processing and innate immune response (**Figure 6c, 6d**), underscoring the impact of the virus on these cells' immune response. Also, a subset of cytokines (CXCL10, CXCL11, CCL8) showed the highest spike of expression in the higher viral titer sub-group (**Supp. Figure 8**), matching previous results from animal models and infected cells (Blanco-Melo *et al*., 2020).

Down-regulated genes and those with a negative enrichment score (NES) were functionally distinct (**Figure 6d**). These included ALAS2, a gene which makes erythroid ALA-synthase that is found in developing erythroblasts (red blood cells). ALA-synthase plays an important role in the production of heme TRIM2 E3 ubiquitin ligase induced during late erythropoiesis, which indicated a connection to hematological and iron (heme) regulation during infection (**Figure 6d**). Accordingly, genes in a related biological network were significantly enriched based on Gene Ontology (GO) pathways for iron regulation (q-value =0.002, **Supp. Table 4**). Both the up-

regulated and down-regulated gene expression differences in were distinct from those of house-keeping genes (**Supp. Figure 9**), which stayed mostly stable during infection.

ACEI usage increases the risk of severe COVID-19 disease

Given our observation of increased *ACE2* gene expression in patients with high SARS-Cov-2 viral load, we broadly investigated the interplay of receiving pharmacologic angiotensin converting enzyme inhibition (ACEI) for hypertension and clinical features of COVID-19 disease. Since *ACE2* expression can be increased in patients taking ACEIs, the observed correlation of viral titer with *ACE2* expression may be attributed to the pre-infection use of such inhibitors, which is common in older patients and those with comorbidities (Fang *et al*., 2020). To address this, we analyzed an observational cohort of 8,278 patients with suspected SARS-CoV-2 infection from NYP were analyzed for their usage of ACEIs (4,574 who tested positive). We found that use of ACEIs was strongly associated with testing positive in patients suspected of SARS-CoV-2 infection (OR=3.06 95% CI: 2.38-3.94, p=2.44E^{-18}). This result was also consistent when corrected for age, sex, and IL-6, where exposure to ACEIs has an OR=1.54, and when corrected for other reported clinical covariates (1.18-2.01, p=1.41E^{-03}); patients exposed to ACEIs were 1.85 (1.40-2.44,p=1.71E^{-05}) times more likely to test positive. In the univariate analysis, ACEI usage conferred an increased risk of intubation and mortality for SARS-CoV-2 positive patients (**Table 1**) (intubation: HR=2.63 95%CI: 2.01-3.43, p=1.22E^{-12})(**Figure 6e**) and mortality: HR=1.68 95%CI: 1.22-2.31, p=1.42E^{-03} (**Figure 6f**).

In a multivariate analysis that included age, sex, and IL-6 levels, a significant association between exposure to ACEIs and those requiring mechanical respiration was also found (HR=1.83 95%CI: 1.39-2.40, p=1.27E^{-05}). Additionally, this association held when correcting for previously reported risk factors for SARS-CoV-2 morbidity and mortality with an HR=1.56 95%CI: 1.18-2.07, p=1.83E^{-03} (**Supp. Figure 10a**). Moreover, we confirmed previously reported risk factors for both mechanical respiration and mortality. For requirement of mechanical respiration, we found significant effects from male sex (p=6.18E^{-03}), diabetes (p=1.00E^{-03}), and IL-6 (p=5.85E^{-113}). In addition, for mortality we found significant effects from age (p=8.05E^{-72}), male sex (p=6.13E^{-04}), diabetes (p=6.66E^{-05}), and IL-6 (p=1.26E^{-12}). Finally, in a post-hoc comparative analysis between specific ACEIs we found that benazepril was associated with significantly increased risk of mortality (N=32, HR=2.37 95%CI: 1.05-5.35, p=3.70E^{-02}) and that trandolapril is associated with requiring mechanical respiration (N=2, HR=15.85 95%CI: 2.11-119.08, p=7.25E^{-03}; **Supp. Figure 10b**).

Discussion

These data demonstrate unique host transcriptome profiles during viral infection with SARS-CoV-2, as well as clade-specific mutational landscapes of the virus, both representing the early phase of the US pandemic (March 2020). Total RNA-seq (ribo-depleted RNA) used here enabled a comprehensive molecular map for the virus and host, including complete genotypes (including insertions and deletions) across the length of the viral genome and coding and non-coding RNA gene expression quantification. Moreover, these data show that LAMP is a quick and effective assay that could be readily deployed to aid current and future viral testing and surveillance, since it is comparable to RNA-seq and RT-PCR and can be used with saliva whole cell lysate. Widespread testing is critical to help inform individual quarantine efforts and overall

management of highly infectious and long-shedding viruses such as SARS-CoV-2, and these data, methods, and results can help guide future clinical and research efforts.

These RNA-sequencing data enabled a genetic map of the viruses in the patients. The phylogenetic analysis provides evidence multiple strains of European origin, but also some (less frequent) strains of Asian origin, confirming recent hypotheses (Gonzalez-Reiche *et al.*, 2020). While NY cases evidence a wide distribution across all identified clades, enrichment of these cases within a subclade of clade A2a (defined by 25563G>T) is suggestive of either multiple founder events, very early introduction, or disproportionate community transmission of strains within this subclade. More broadly, these findings may indicate multiple introductions of SARS-CoV-2 to the New York City area, likely by people travelling from affected regions in Europe. The existence of this subclade within a clade of putative European origin suggests that founder effects of early transmissions or differences in virulence between strains of European and Asian origin play a significant role in the ongoing dynamics of transmission within New York City. (**Figure 4)**. Because genomes provided here were recovered from patients that arrived at the hospital, these data may only partly reflect the strains present in asymptomatic patients. Indeed, the continuous production of viral genomes in sick and healthy carriers can be used to track the pathogenicity and biology of SARS-CoV-2 strains as they move around the world, especially when paired with clinical and outcome data. Moreover, further transcriptomic and epitranscriptomic analysis and mapping through these sequencing efforts will provide new insights to the life cycle and pathogenicity of SARS-CoV-2 (Kim et al., 2020).

During a large-scale pandemic, with exponentially infectious spread as seen with COVID-19, scalable methods for diagnosis and screening of pathogens that can build the capacity to test millions of patients are desperately needed (Lan *et al*., 2020, Liu *et al*., 2020). As such, LAMP and CRISPR-based methods provide an opening for more rapid screening options, and more importantly a point-of-care test (POCT) that can translate into clinical workflow. Moreover, the data presented here represent the first full clinical validation of such rapid-response, nucleic acid-based diagnostic approaches, matched with a standard diagnostic method (RT-PCR) and total RNA-seq. The limitations of the sampling collection (e.g. swab type, swabbing method) as well as the detection technology will help inform how the results should be interpreted, and how some of the false negative results on LAMP can occur for simple technical reasons. For example, the available primers used for the clinical specimens on RT-PCR positive results can indicate if an area is infectious and a negative result (with appropriate confirmation) will possibly represent a lower risk. Indeed, these tools and methods can help create a viral "weather report" if broadly used and if partnered with continual validation.

Although much of the pathophysiology of the novel coronavirus, SARS-CoV-2, remains unknown, studies have reported the virus using the receptor ACE2 for entry into target cells (Hoffman *et al*., 2020; Lu *et al*., 2020). Further validating this, we found expression of the *ACE2* gene was significantly upregulated in the samples with higher viral load, as well as some cytokine genes, which matches observations found in other betacoronaviruses (Sajuthi *et al*., 2020). ACE inhibitors are commonly used in COVID-19 patients with comorbidities of hypertension, diabetes mellitus, and coronary heart diseases (Fang *et al*., 2020, Ferrario et al., 2005), concerning physicians that it may make patients on these medications more susceptible to SARS-CoV-2 infection. Indeed, since epidemiological studies have reported increased mortality

and morbidity in COVID-19 patients with hypertension, more research needs to be done to address other confounding variables including other comorbidities and treatment with ACE inhibitors or angiotensin receptor blockers (Patel and Verma, 2020).

Although these data alone cannot establish causality between infection and ACE2 expression regulation, they provide some testable hypotheses. For example, if some patients are more susceptible because they are already expressing high levels of ACE2, this could help with targeting ACE2 in these patients as a prophylactic method. However, if the cells respond to infection with ACE2 expression, and this leads to the cytokine storm seen in patients, then this could be used as a downstream treatment (post-infection), for when ACE2 interacts with TMPRSS2, such as the ongoing trials with camostat mesylate (Hoffman *et al*., 2020). Moreover, SARS-CoV-2 has been found to induce phosphorylation of STAT1 and increases in interferon stimulated gene proteins, a mechanism not previously seen in SARS-CoV, suggesting a potential molecular mechanism behind the upregulation of interferon response (Lokugamage et al., 2020).

Finally, it is notable that the majority of the testing for SARS-CoV-2 so far has relied on nasopharyngeal specimen collection, yet preliminary results here and elsewhere demonstrate that COVID-19 prediction from oral collection could be a more optimal path forward (Woelfel *et al*., 2020). However, further studies comparing nasopharyngeal, oropharyngeal, and buccal collection approaches, as well as a comparison of different swab types, are needed. Depending on the availability of reagents and resources, as well as automation, a LAMP-based approach on such sample types could allow facilities to increase testing capabilities by orders of magnitude. Since viral pandemics can have significant, long-lasting detrimental impacts for affected countries, it is crucial to deploy methods that can track and profile cases (e.g. RNA-seq, LAMP, RT-PCR) and provide a comprehensive view of host and viral biology. These methods can help mitigate the medical and socioeconomic harm from viral outbreaks, as well as establish protective surveillance networks that can help defend against future outbreaks.

Acknowledgements
We thank the Core Facilities at Weill Cornell Medicine, the Scientific Computing Unit (SCU), and the XSEDE Supercomputing Resources. We also thank New England Biolabs for providing the reagents for preliminary testing of the LAMP protocols, as well as Eileen Dimalanta, and Ted Davis, for technical discussions. The authors wish to thank the following members of the HudsonAlpha Discovery team who supported the RNASeq experiments described in the manuscript: Colleen Cowan, John Mote, Arianna Pionzio, Melanie Robinson, and Madison Robison.

Funding
We are grateful for support from the STARR Foundation (I13-0052) the Vallee Foundation, the WorldQuant Foundation, The Pershing Square Sohn Cancer Research Alliance, Citadel, the National Institutes of Health (R01MH117406, R25EB020393, R01AI151059), the Bill and Melinda Gates Foundation (OPP1151054), the NSF (1840275), the National Center for Advancing Translational Sciences of the National Institutes of Health (UL1TR000457, TL1TR0002386-03, CTSC), and the Alfred P. Sloan Foundation (G-2015-13964). FJS would

like to thank the National Institute of Allergy and Infectious Diseases (1U19AI144297-01). MZ supported by T15LM007079. NPT and UOG supported by R35GM131905.

Author Contributions
CEM led the study design and coordination and HR led the clinical collection and validation work in the NYP CLIA laboratory, as well as with MI. Overall supervision and protocol development and implementation for the Zymo RNAClean and NEB assays (SL). DB and ChM performed the LAMP experiments to validate the assay. CeM, DD, JF, AS, JR, MM, EA, IH, DM, MI, BWL, MZ, UG, NiT, NI, CEM performed analyses. DD, NI, MS, BY, KR, CB coordinated and collected environmental samples. EA submitted the IRB application and helped with clinical coordination. Help and insights for analysis from EA, SL, MI, MS, LFW, ML, MC, HR all led to the figures and analyses. All authors reviewed, edited, and approved the manuscript.

Conflicts of Interest
Nathan Tanner and Brad Langhorst are employees at New England Biolabs.

IRB
Samples were collected and processed through the Weill Cornell Medicine Institutional Review Board (IRB) Protocol 19-11021069. Observational cohort analysis (ACEI) was done through the Columbia University IRB Protocol AAAL0601.

Data Accessibility
Viral sequences were uploaded into GISAID (Global Initiative on Sharing All Influenza Data) site (https://www.gisaid.org), and patient data is deposited into dbGAP.

Materials and Methods
Sample Collection and Processing
Patient specimens were collected with patients' consent at New York Presbyterian Hospital (NYPH) and then processed for RT-PCR. Nasopharyngeal (NP) swab specimens were collected using the BD Universal Viral Transport Media system (Becton, Dickinson and Company, Franklin Lakes, NJ) from symptomatic patients.

Extraction of Viral RNA and RT-PCR detection
Total viral RNA was extracted from deactivated samples using automated nucleic acid extraction on the QIAsymphony and the DSP Virus/Pathogen Mini Kit (QIAGEN). One step reverse transcription to cDNA and real-time PCR (RT-PCR) amplification of viral targets, E (envelope) and S (spike) genes and internal control, was performed using the Rotor-Gene Q thermocyler (QIAGEN).

Twist Synthetic RNAs
We used two fully synthetic RNAs made by in vitro transcription (IVT) from Twist Biosciences, which were synthesized in 5kb pieces with full viral genome coverage of SARS-CoV-2. They were sequence verified to ensure >99.9% viral genome coverage, and come as 1,000,000 copies per µL, 100µL per tube. The two controls are from Wuhan, China (MN908947.3) and

Melbourne, Australia (MT007544.1). Reference sequence designs came from NCBI: https://www.ncbi.nlm.nih.gov/nuccore/MT007544 and https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3.

Reverse Transcriptase, quantitative real-time PCR (RT-PCR)
Clinical samples were extracted as described above and then tested with RT-PCR using primers for the E (envelope) gene, which detects all members of the lineage B of the beta-CoVs, including all SARS, SARS-like, and SARS-related viruses, and a second primer set for the S (spike) gene, which specifically detects the SARS-CoV-2 virus. The reaction also contains an internal control that served as an extraction control and a control for PCR inhibition.

Samples were annotated using RT-PCR cycle threshold (Ct) value for SARS-CoV-2 primers. Subjects with Ct less than or equal to 18 were assigned "high viral load" label, Ct between 18 and 24 were assigned "medium viral load" and Ct between 24 and 40 were assigned "low viral load" classes, with anything above Ct of 40 was classified as negative. We also predicted a combined viral load score using Ct, GloMax QuantiFluor readout from LAMP experiments and fraction of SARS-CoV-2 matching NGS reads in a sample. For this score (40-Ct), (LAMP readout) and (log10(SARS-CoV-2 fraction + 1e-6)) were all normalized between zero and one individually, and summed together using a combination weight of 5 for Ct, 3 for LAMP and 2 for NGS.

LAMP Primer Sequences
Primers were designed using PrimerExplorer (v4.0), as per guidelines in Zhang *et al*., 2020. This specifically utilized the LAMP-compatible primers for the on the COVID-19 reference genome (NCBI). LAMP's inherent specificity (using 4-6 primers vs. 2 for PCR amplification) in combination with this *in-silico* analysis revealed there is limited opportunity for cross-reactivity to allow for false-positive reporting or affect performance of the N-gene primers for SARS-CoV-2 detection (**Supp Table 4**). Overall, the primers had less than 80% homology with the vast majority of tested pathogen sequences. For any organisms where a primer hit >80% homology, only one of the primers (forward or reverse) had significant homology making an amplified product extremely unlikely. Overall, the results of this analysis predict no significant cross-reactivity or microbial interference. We also assessed the potential impact of sequence variation in circulating strains that might lead to poor amplification. In the thousands of sequences deposited in GISAID (Shu and McCauley, 2017), only one site in the priming region was observed to be polymorphic (see Supplemental Figure X+1). The polymorphism (T30359C) was only observed in 106 of 6753 (<2%) sequences with coverage of this region. This variant overlaps the priming site of the LB primer but is not near a 3-prime end and is not anticipated to cause amplification failure.

Nucleocapsid Gene: (N, 5'-3'):
GeneN-F3 TGGCTACTACCGAAGAGCT
GeneN-B3 TGCAGCATTGTTAGCAGGAT
GeneN-FIP TCTGGCCCAGTTCCTAGGTAGTCCAGACGAATTCGTGGTGG
GeneN-BIP AGACGGCATCATATGGGTTGCACGGGTGCCAATGTGATCT
GeneN-LF GGACTGAGATCTTTCATTTTACCGT
GeneN-LB ACTGAGGGAGCCTTGAATACA

Primer Sequences: rActin (5′-3′)

ACTB-F3	AGTACCCCATCGAGCACG
ACTB-B3	AGCCTGGATAGCAACGTACA
ACTB-FIP	GAGCCACACGCAGCTCATTGTATCACCAACTGGGACGACA
ACTB-BIP	CTGAACCCCAAGGCCAACCGGCTGGGGTGTTGAAGGTC
ACTB-LoopF	TGTGGTGCCAGATTTTCTCCA
ACTB-LoopB	CGAGAAGATGACCCAGATCATGT

The LAMP Reaction Setup

For each well or Eppendorf tube, we utilized a set of six primers (above) for Gene N, the M1800 LAMP Master Mix (NEB), water, and 11.5µL of the sample. The protocol is as follows:

1. Reagents added:
    a. 12.5 µL M1800 LAMP mix (NEB)
    b. 1 µL LAMP primers (Gene N)
    c. 1-11.5 µL of sample
    d. Remaining volume (to 25 µl) H_2O
2. Vortex, spin down;
3. Place on Thermocycler at 65ºC for 30 minutes with lid at 105 ºC;
4. Remove tubes, place on ice for 5 seconds;
5. Visualize over lab bench/ice/paper.

Light Intensity and Data Processing

Completed reactions were analyzed with the Promega GloMax Explorer (Promega GM3500) fluorometer using the QuantiFluor ONE dsDNA system (Promega E4871). This system recorded light intensity from each well using an emission filter of 500-550nm, an excitation filter set at blue 475 nm, and a high sensitivity setting on the Glomax software. Values were then tabulated and compared with controls (positive and negative). The intensity threshold of 2.5x negative control was used as the threshold for positive detection.

DNAse treatment, rRNA depletion, and RNAseq library construction

All samples' total nucleic acid (TNA) were treated with DNAse 1(Zymo Research, Catalog # E1010), which cuts both double-stranded and single-stranded DNA. Post-DNAse digested samples were then put into the NEBNext rRNA depletion v2 (Human/Mouse/Rat), Ultra II Directional RNA (10ng), and Unique Dual Index Primer Pairs were used following the vendor protocols from New England Biolabs (except for the first flowcell, see supplelmental figures). Kits were supplied from a single manufacturer lot. Completed libraries were quantified by Qubit or equivalent and run on a Bioanalyzer or equivalent for size determination. Libraries were pooled and sent to the WCM Genomics Core, HudsonAlpha, or New York Genome Center for final quantification by Qubit fluorometer (ThermoFisher Scientific), TapeStation 2200 (Agilent), and RT-qPCR using the Kapa Biosystems Illumina library quantification kit.

Taxonomic Classification of Sequence Data

All complete genome or chromosome level assemblies from RefSeq database for archaea, bacteria, protozoa, fungi, human and viruses including SARS-CoV and SARS-CoV-2 genomes

were downloaded and used for building a classification database for Kraken2 (k=35, ℓ=31) (O'Leary *et al.,* 2016; Wood *et al.,* 2019).

To get an approximation for the positive and negative classification rate, the BBMap random-reads script was used to simulate 10 million 150bp paired-end Illumina reads from the database sequences (Segata *et al.,* 2016). For the negative test all sequences in the database excluding SARS-CoV and SARS-CoV-2 genome were removed from the sequences and the simulated reads were mapped with the Kraken2 database (Supp Table 1).

For the positive test, the same process was repeated using only SARS-CoV-2 genome (**Supp Table 1**). Positive results show >99% of SARS-CoV-2 reads uniquely map to either SARS-CoV or SARS-CoV-2, with the remaining 1% are ambiguous, potentially matching multiple taxa (**Supp Table 2**). All sequences were classified using the Kraken2 database. To remove the potential contamination of reads that are homologous across multiple species we used Kraken2 outputs to filter sequences to either human (uniquely matching Homo sapiens and no other taxon in our database), SARS-CoV-2 (either matching SARS-CoV or SARS-CoV-2 due to homology between these two viruses), and remaining reads that may be coming from unclassified, archaeal, bacterial, viral, fungal, protozoan or ambiguously mapping reads to human or SARS-CoV (*Li*, 2015).

Reference- and Fragment-based Viral Mapping and Assembly
Reads unambiguously mapping to SARS-CoV or SARS-CoV-2 were aligned to the Wuhan-Hu-1 (Genbank accession MN908947.3) reference using bwa mem (Li, 2013). Variants were called using iVar, and pileups and consensus sequences were generated using samtools (Li *et al.,* 2009; Grubaugh *et al.,* 2019; Greenfield *et al.,* 2020). Any sample with more than 30,000 SARS-CoV-2 mapping reads and >99% coverage above 10x depth were taken as reliable samples, which resulted in 95 samples (90 positive, 2 negative, 2 positive controls and 1 negative control). 92 clinical samples were compared to 4872 SARS-CoV-2 sequences from GISAID (as of April 10, 2020) (9, 10). All sequence filtering, alignments, phylogenetic inference, temporal ordering of sequences and geographic reconstruction of likely transmission events were done using nextstrain (Katoh and Standley, 2013; Sagulenko *et al.,* 2018; Hadfield *et al.,* 2018).

Fragment assembly was also performed using IRMA (Sehpard *et al.,* 2016). The Wuhan-Hu-1 genome (Genbank accession MN908047.3) was used as a reference with the poly-A tail trimmed to reduce the likelihood of false low-complexity matching. An HMM model of this reference sequence was created using the native modelfromalign script. IRMA was run with the COV module with the following parameters adjusted from the default: minimum read length = 25bp; minimum read pattern count = 5; minimum read count = 2; minimum count for alternative finished assembly = 20. Consensus assemblies were aligned to the Wuhan-Hu-1 reference using MAFFT (Katoh and Standley, 2013) with default settings, and sequence identity and coverage metrics were calculated using Mview (Brown *et al.,* 1998). Phylogenetic trees were created using nextstrain's augur as described above, and visualized using the ggtree package in R (Yu, 2020).

Viral SNV Calling
Subsequent to alignment, SNV were identified by xAtlas (Farek *et al.,* 2018) (parameters: -g -m 10 -n 10) producing a gVCF per sample. The individual call sets were merged using bcftools (Li,

2011) (merge -g) and allele counts were added with bcftools view (-q 0.01) . Population-level variants across all samples were converted in a presence/absence matrix using a customized script where every SNV was counted as present if it was supported by at least 5% of the reads. This matrix was loaded in R and underwent hierarchical clustering using the function dist with the euclidean method and hclust using the average method. All samples were further analyzed using Manta (Chen *et al.,* 2016), and subsequently summarized and filtered using the SURVIVOR pipeline (Jeffares *et al.,* 2017). Subsequently, SV calls were re-genotyped using SVTyper (doi:10.1038/nmeth.3505) and a population-level VCF was generated using SURIVOR merge (allowing for 100bp maximum distance). Events were filtered for only <1000bp insertions and deletions that were homozygous in a sample.

Human Transcriptome Analysis

The reads that mapped unambiguously to the human reference genome via Kraken2 were used to detect the host transcriptional response to the virus. Reads matching *Homo sapiens* were trimmed with TrimGalore, aligned with STAR (v2.6.1d) to the human reference build GRCh38 and the GENCODE v33 transcriptome reference, gene expression was quantified using featureCounts, stringTie and salmon using the nf-core RNAseq pipeline (Pertea *et al.,* 2015; Malinen *et al.,* 2005; Johnson *et al.,* 2007; Robinson *et al.,* 2010; Naccache *et al.,* 2014; Zamani *et al.,* 2017; Ewels *et al.,* 2019). Sample QC was reported using fastqc, RSeQC, qualimap, dupradar, Preseq and MultiQC (Okonechnikov *et al.,* 2016; Andrews, 2015; Ewesl *et al.,* 2016; Sayols *et al.,* 2016; Wang *et al.,* 2012). Samples that had more than 10 million human mapped reads were used for differential expression analysis. Reads, as reported by featureCounts, were normalized using variance-stabilizing transform (vst) in DESeq2 package in R (Love *et al.,* 2014) for visualization purposes in log-scale. DESeq2 was used to call differential expression with either Positive cases vs Negative, or viral load (High/Medium/Low/None) as reported by either RT-PCR cycle threshold (Ct) values or the combination viral load method as explained before. Genes with BH-adjusted p-value < 0.01 and absolute log2 fold-change greater than 0.58 (at least 50% change in either direction) were taken as significantly differentially regulated (Benjamini and Hochberg, 1995). The complete gene list for all comparisons are given in Supp Table 3. Resulting gene sets were ranked using log2 fold-change values within each comparison and put into GSEA to calculate gene set enrichment for molecular signatures database (MSigDB), MGI Mammalian Phenotypes database and ENCODE transcription factor binding sets (Liberzon *et al.,* 2011; Subramanian *et al.,* 2005; Sergushichev, 2016; Smith *et al.,* 2018). Any signature with adjusted p-value < 0.01 and absolute normalized enrichment score (NES) >= 1.5 were reported (Supp Table 3).

Statistics and Visualization

Statistics and visualization of single nucleotide variants and indels were completed in R. Visualization of phylogenies was completed using Auspice and the 'ape' library for R.

Cross-reactivity Analysis

Primers were compared with a list of sequences from organism from the same genetic family as SARS-CoV-2 and other high-priority organisms listed in the United States Food and Drug Administration's Emergency Use Authorization Template (https://www.fda.gov/media/135900/download). Using the sequence names in the EUA template, the NCBI taxonomy database was queried to find the highest quality representative sequences for

more detailed analysis. Primers were compared to this database using Blast 2.8.1 and the following parameters (word size: 7, match score: 2, mismatch score: -3, gap open cost: 5, gap extend cost: 2). Up to 1000 hits with e-value > 10 were reported. Supplemental Figure

Inclusivity Analysis
Unique, full-length, human-sample sequences were downloaded from the GISAID web interface. These sequences were aligned to NC_045512.2 (Wuhan SARS-CoV-2) using minimap2 -x asm5 and visually inspected using IGV 2.8.0 with allele frequency threshold set to 0.01.

ACE Inhibitor Cohort Analysis
We compared usage of ACE inhibitors in an observational cohort analysis of 8,278 patients with suspected SARS-CoV-2 infection (4,574 of which tested positive). ACE inhibitors are commonly taken continuously for several years (Bonarjee et al., 2001). We defined a cohort of ACE inhibitor-exposed patients as those that have an ACE inhibitor prescription order sometime after January 1st, 2019. We compared the frequency of ACE inhibitor exposure in three cohort comparisons:
i. SARS-CoV-2 tested positive patients versus SARS-CoV-2 tested negative patients,
ii. SARS-CoV-2 positive patients who require mechanical ventilation versus those who did not, and
iii. SARS-CoV-2 positive patient survival versus death.

In addition, we perform one post-hoc comparison to evaluate the individual effects of particular ACE inhibitors among SARS-CoV-2 positive patients who are exposed to ACE inhibitors.

Cohort and Data Source
Our cohort data for SARS-CoV-2 suspected patients is extracted from the electronic health records at New York-Presbyterian/Columbia University Irving Medical Center (NYP/CUIMC). We used data collected starting on March 10th, 2020 through April 16th, 2020. In addition, we used data from 279,487 patients, who were not tested for SARS-CoV-2 infection, with available electronic health records from January 1st, 2019 through September 24th, 2019 to represent a comparison population of patients. In both cases, data extracted included disease diagnoses, laboratory measurements, medication and pharmacy orders, and patient demographics. We derived mortality from a death note filed by a resident or primary provider that records the date and time of death. Intubation was used as an intermediary endpoint and is a proxy for a patient requiring mechanical respiration. We used note types that were developed for patients with SARS-CoV-2 infection to record that this procedure was completed. We validated outcome data derived from notes against the patient's medical record using manual review.

Experimental and Statistical Methods
We conducted univariate analysis of the frequency differences of ACE inhibitor exposure and multivariate regression analysis to account for established risk factors for SARS-CoV-2 outcomes (i.e. age, sex, and baseline IL-6 upon admission (Zhou et al., 2020). We use logistic regression for analysis for comparing cohort (i) as this cohort is retrospective analysis on patients with definitive outcomes. We perform survival analysis using a Cox proportional hazards model for cohort comparisons (ii) and (iii) as some patients from these cohorts are currently being treated and their outcomes are unknown (i.e., censored). We cannot determine from our data the

date of infection. For the study start date for the patient, we use the date of testing positive minus seven days. Quantitative variables (i.e., age and the first IL-6 measurement) are scaled to [0,1] to facilitate comparison of model coefficients to those of dichotomous variables. Prior to conducting our multivariate analysis, we evaluated and removed correlated covariates. We include a model built with all covariates as well in the supplemental tables.

Comorbidity Definitions
Risk factors were assigned using OMOP CDM concept IDs 317576 for "Coronary Arteriosclerosis", 201820 for "Diabetes mellitus", and 437525 for "Overweight". In each case, the concept ID and all descendant concepts were used to define the risk factor phenotype, and individuals were assigned the phenotype if they were assigned any of the codes.

Statistical Software
All analyses were done in using Python 3.7 and all models were fit using R 3.6.3. Survival analyses (Cox regressions and survival curves) were performed with the survival package for R, version 3.1-12.

References

Amanat F, Nguyen T, Chromikova V, Strohmeier S, Stadlbauer D, Javier A, Jiang K, Asthagiri-Arunkumar G, Polanco J, Bermudez-Gonzalez M, Caplivski D, Cheng A, Kedzierska K, Vapalahti O, Hepojoki J, Simon V, Krammer F. A serological assay to detect SARS-CoV-2 seroconversion in humans. *MedRxiv. 2020 Mar 18.* https://www.medrxiv.org/content/10.1101/2020.03.17.20037713v1

Andrews S. FASTQC A Quality Control tool for High Throughput Sequence Data. *Babraham Inst.* (2015).

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* (1995)

Broughton JP, Deng X, Yu G, Fasching CL, Singh J, Streithorst J, Granados A, Sotomayor-Gonzalez A, Zorn K, Gopez A, Hsu E, Gu W, Miller S, Pan C, Guevara H, Wadford D, Chen J, Chiu CY. Rapid Detection of 2019 Novel Coronavirus SARS-CoV-2 Using a CRISPR-based DETECTR Lateral Flow Assay. 2020 Mar; 10: 1-27. https://medrxiv.org/cgi/content/short/2020.03.06.20032334v1

Brown, N.P., Leroy C., Sander C. (1998). MView: A Web compatible database search or multiple alignment viewer. *Bioinformatics*. 14 (4):380-381.

Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S, Saunders CT. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications, *Bioinformatics*, Volume 32, Issue 8, 15 April 2016, Pages 1220–1222

Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, Marth GT, Quinlan AR, Hall IM. SpeedSeq: Ultra-fast personal genome analysis and interpretation. *Nature Methods* 12(2015): 966–968.

Danko D, Bezdan D, Afshinnekoo E, et al…Mason CE. Global Genetic Cartography of Urban Metagenomes and Anti-Microbrial Resistance https://www.biorxiv.org/content/10.1101/724526v1

Deng X, Achari A, Federman S, Yu G, Somasekar S, et al…Chiu CY. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nature Microbiology.* 2020 Mar; 5(3): 443–454.

Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* 2020 Feb 4.

Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Challenges* (2017)

Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* (2016)

Ewels PA et al., nf-core: Community curated bioinformatics pipelines. *Nature* 2020 Feb 13.

Farek J, Hughes D, Mansfield A, Krasheninina O, Nasser W, Sedlazeck FJ, Khan Z, Venner E, Metcalf G, Boerwinkle E, Muzny DM, Gibbs RA, Salerno W. xAtlas: Scalable small variant calling across heterogeneous next-generation sequencing experiments. bioRxiv 295071; doi: https://doi.org/10.1101/295071

Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir Med. 2020 Apr;8(4):e21.

Fauver JR, et al., Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell* 2020.

Gardy J, Loman NJ, Rambaut A. Real-time digital pathogen surveillance - the time is now. Genome Biology. 2015 Jul 30;16(1):155.

Gonzalez-Reiche A, Hernandez MH, Sullivan M, Ciferri B, Alshammary H, Obla A, Fabre S, Kleiner G, Polanco J, Khan Z, Alburquerque B, Guchte A, Dutta J, Francoeur N, Salom B, Oussenko I, Deikus G, Soto J, Sridhar SH, Wang Y, Twyman K, Kasarskis A, Altman DR, Smith M, Sebra R, Aberg J, Krammer F, Garcia-Sarstre A, Luksza M, Patel G, Paniz-Mondolfi A, Gitman M, Sordillo EM, Simon V, Bakel H. Introductions and early spread of SARS-CoV-2 in the New York City area. MedAxiv. 2020. Apr 11. https://www.medrxiv.org/content/10.1101/2020.04.08.20056929v1

Guan WJ, Ni Zy, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Zhong NS, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. *NEJM*. 2020 Feb 28.

Greenfield N, SARS-CoV-2 variant calling and consensus assembly pipeline. GitHub (2020), (available at https://github.com/onecodex/sars-cov-2/).

Grubaugh ND et al., An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol*. (2019)

Hadfield J, Megill C, Bell S, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, and Neher RA. 2018. Nextstrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics* 34 (23): 4121–23.

He F, Deng Y, Li W. Coronavirus Disease 2019 (COVID-19): What we know? J Med Virol. 2020 Mar 14.

Hoffmann M, Weber HK, Schroeder Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu N, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020. Mar 5.

Jeffares DC, Jolly C, Hoti M, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*. 2017 Jan;8:14061.

Jenkins C and Orsburn B. In silico approach to accelerate the development of mass spectrometry-based proteomics methods for detection of viral proteins: Application to COVID-19. 2020 Mar 10. https://www.biorxiv.org/content/10.1101/2020.03.08.980383v1

Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* (2007).

Konrad R, Eberle U, Dangel A, Treis B, Berger A, Bengs K, Fingerle V, Liebl B, Ackermann N, Sing A. Rapid Establishment of Laboratory Diagnostics for the Novel Coronavirus SARS-CoV-2 in Bavaria, Germany, February 2020. *Eurosurveillance*. 2020 Mar 5.

Kim D, Lee J, Yang J, Kim JW, Kim N, Chang H. The architecture of SARS-CoV-2 transcriptome. *BioRxiv* https://www.biorxiv.org/content/10.1101/2020.03.12.988865v2

Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, Xu H. Positive RT-PCR Test Results in Patients Recovered from COVID-19. *JAMA* 2020 Feb 27.

Langmead B, and Salzberg S. 2012. Fast Gapped-Read Alignment with Bowtie 2. *Nature Methods* 9 (4): 357–59.

Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*. 2020 Mar 10. M20-0504.

Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* 2014;15(6):R84.

Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987–2993.

Li H, Seqtk: Toolkit for processing sequences in FASTA/Q formats. GitHub (2015).

Li H, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Prepr. arXiv (2013), doi:arXiv:1303.3997 [q-bio.GN].

Li H et al., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009).

Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*. 2020 Mar 16.

Liberzon A et al., Molecular signatures database (MSigDB) 3.0. *Bioinformatics* (2011).

Liu R, Han H, Liu F, Lv Z, Wu K, Liu Y, Feng Y, Zhu C. Positive Rate of RT-PCR Detection of SARS-COV-2 Infection of 4880 Cases from One Hospital in Wuhan, China, from Jan to Feb 2020. *Clinica Chimica Acta* 505 (2020) 172-175.

Lokugamag KG, Hage A, Schindewolf C, Rajsbaum R, Menachery VD. SARS-CoV-2 is sensitive to type 1 interferon pretreatment. *BioRxiv* https://www.biorxiv.org/content/10.1101/2020.03.07.982264v3

Love MI et al., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014), doi:10.1186/s13059-014-0550-8.

Lu J et al., Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* 2020

Malinen E et al., Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR. *Am. J. Gastroenterol*. 100, 373–382 (2005).

Metsky HC, Freije CA, Kosoko-Thoroddsen TF, Sabeti PC, Myhrvold C. CRISPR-based surveillance for COVID-19 using genomically-comprehensive machine learning design. 2020 Mar 2. *BioRxiv*. https://www.biorxiv.org/content/10.1101/2020.02.26.967026v2.

Meyers L, Dien Bard J, Galvin B, Nawrocki J, Niesters HGM, Stellrecht KA, St George K, Daly JA, Blaschke AJ, Robinson C, Wang H, Cook CV, Hassan F, Dominguez SR, Pretty K, Naccache S, Olin KE, Althouse BM, Jones JD, Ginocchio CC, Poritz MA, Leber A, Selvarangan R. Enterovirus D68 Outbreak Detection Through A Syndromic Disease Epidemiology Network. *J Clin Virol*, 124, 104262, 2020 Mar.

Naccache SN et al., A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.* (2014),

Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* (2016),

Ong SWX, Tan YK, Chia PY, Lee TH, Ng OT, Wong MSY, Marimuthu K. Air, Surface Environmental, and Personal Protective Equipment Contamination by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) From a Symptomatic Patient. *JAMA*. 2020 Mar 4.

O'Leary NA et al., Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. (2016).

Patel AB, Verma A. COVID-19 and Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers. What is the evidence? *JAMA* 2020 Mar 24.

Peci A, Winter AL, King EC, Blair J, Gubbay JB. Performance of rapid influenza diagnostic testing in outbreak settings. J Clin Microbiol. 2014 Dec;52(12):4309-17.

Pertea M et al., StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol*. (2015)

Quilty B, Clifford S, Flasche S, Eggo RM. Effectiveness of airport screening at detecting travellers infected with 2019-nCoV. 2020. Feb 02. *MedRxiv*. https://www.medrxiv.org/content/10.1101/2020.01.31.20019265v1

Robinson MD, Oshlack A, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* (2010).

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, Volume 28, Issue 18, 15 September 2012, Pages i333–i339.

Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. 2011. Integrative Genomics Viewer. *Nature Biotechnology* 29 (1): 24–26.

Santarpia JL, Rivera DN, Herrera V, Morwitzer MJ, Creager H, Santarpia GW, Crown KK, Brett-Major DM, Schnaubelt E, Broadhurst MJ, Lawler JV, Reid SP, Lowe JJ.

Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of Nebraska Medical Center. 2020. Mar 26. *MedRxiv*. https://www.medrxiv.org/content/10.1101/2020.03.23.20039446v2

Segata N et al., BBMap short-read aligner, and other bioinformatics tools. *Bioinformatics* (2016).

Sajuthi SP, DeFord P, Jackson ND, Montgomery MT, Everman JL, Rios CL, Pruesse E, Nolin JD, Plender EG, Wechsler ME, Mak A, Eng C, Salazar S, Medina V, Wohlford EM, Huntsman S, Nickerson D, Germer S, Zody MC, Abecasis G, Kang H, Rice KM, Kumar R, Oh S,

Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* (2017).

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* (2013).

Rodriguez-Santana J, Burchard EG, Seibold MA. Type 2 and interferon inflammation strongly regulate SARS-CoV-2 related gene expression in the airway epithelium. Biorxiv. 2020. Apr 10. https://www.biorxiv.org/content/10.1101/2020.04.09.034454v1

Sagulenko P, Puller V, Neher RA, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* (2018), doi:10.1093/ve/vex042.

Sayols S, Scherzinger D, Klein H. dupRadar: A Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics* (2016), doi:10.1186/s12859-016-1276-2.

Sergushichev A, An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv (2016), doi:10.1101/060012.

Shepard, S.S., Meno, S., Bahl, J. et al. Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler. *BMC Genomics* 17, 708 (2016).

Schmid-Burgk JL Li D, Feldman D, Słabicki1 M, Borrajo J, Strecker J, Cleary1 B, Regev A, Zhang F. LAMP-Seq: Population-Scale COVID-19 Diagnostics Using a Compressed Barcode Space. BioRxiv. https://www.biorxiv.org/content/10.1101/2020.04.06.025635v1

Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ, Mouse Genome Database (MGD)-2018: Knowledgebase for the laboratory mouse. *Nucleic Acids Res*. (2018),

Subramanian A et al., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* (2005),

Taiaroa G, Rawlinson D, Featherstone L, Pitt M, Caly L, Druce J, Purcell D, Harty L, Tran T, Roberts J, Catton M, Williamson D, Coin L, Duchene S. Direct RNA sequencing and early evolution 2 of SARS-CoV-2. 2020 Mar 7. *BioRxiv*. https://www.biorxiv.org/content/10.1101/2020.03.05.976167v1

Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J, Lu J. On the origin and continuing evolution of SARS-CoV-2, *National Science Review*. 2020. Mar 3.

Tanner NA, Zhang Y, Evans TC. Visual Detection of Isothermal Nucleic Acid Amplification Using pH-Sensitive Dyes. *Biotechniques*. 58 (2), 59-68. 2015 Feb 1.

van Doremalen N, Bushmaker T, Morris D, Holbrook M, Gamble A, Williamson B, Tamin A, Harcourt J, Thornburg N, Gerber S, Lloyd-Smith J, de Wit E, Munster V. Aerosol and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1. 2020. Mar 17. *New England Journal of Medicine*.

Wang L, Wang S, Li W, RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* (2012)

Woelfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Kelly TCJ, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brünink S, Schneider J, Ehmann R, Zwirglmaier K, Drosten C, Wendtner C. Virological assessment of hospitalized cases of coronavirus disease 2019 . 2020. Mar 8. https://www.medrxiv.org/content/10.1101/2020.03.05.20030502v1

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. (2019).

Yu G. Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics*, 2020, 69:e96. doi: 10.1002/cpbi.96.

Yu L, Wu S, Hao X, Li X, Liu X, Ye S, Han H, Dong X, Li X, Li J, Liu J, Zhang W, Pelechano V, Chen W, Yin X. Rapid Colorimetric Detection of COVID-19 Coronavirus Using a Reverse Transcriptional Loop-Mediated Isothermal Amplification (RT-LAMP) Diagnostic Platform: iLACO. 2020 Feb 24. https://www.medrxiv.org/content/10.1101/2020.02.20.20025874v1

Zamani S et al., *Mycobacterium avium* subsp. *paratuberculosis* and associated risk factors for inflammatory bowel disease in Iranian patients. *Gut Pathog*. 9, 1 (2017).

Zhao Z, Sokhansanj BA, Rosen G. Characterizing geographical and temporal dynamics of novel coronavirus SARS-CoV-2 using informative subytype markers. *BioRxiv* https://www.biorxiv.org/content/10.1101/2020.04.07.030759v3

Zhang Y, Odiwuor N, Xiong J, Sun L, Nyaruaba RO, Wei H, Tanner NA. Rapid Molecular Detection of SARS-CoV-2 (COVID-19) Virus RNA Using Colorimetric LAMP. 2020. Feb 29. *MedRxiv* https://www.medrxiv.org/content/10.1101/2020.02.26.20028373v1

Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of

adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020 Mar 11.

Figures and Legends

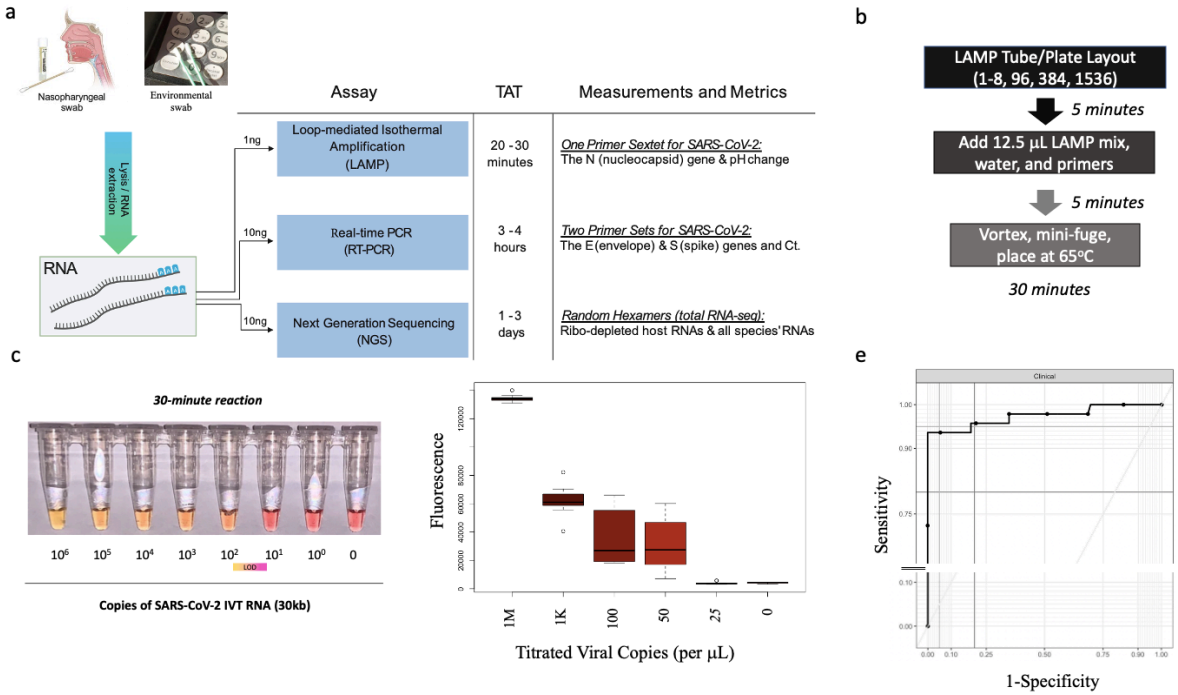


Figure 1. Sample Processing, the Loop-Mediated Isothermal (LAMP) Reaction and Synthetic RNA Validation. (A) Clinical and environmental samples collected with Nasopharyngeal (NP) and isohelix swabs respectively, were tested with RNA-sequencing, RT-PCR, and LAMP. (B) The test samples were prepared using an optimized LAMP protocol from NEB, with a reaction time of 30 minutes. (C) Reaction progress was measured for the Twist COVID-19 synthetic RNA (*MT007544.1*) from 1 million molecules of virus (10^6), then titrated down by $\log_{10}$ dilutions. The colorimetric findings of the LAMP assay are based on a yellow to pink gradient with higher copies of SARS-CoV-2 RNA corresponding to a yellow color. The limit of detection (LoD) range is shown with a gradient after 30 minutes between 10 and 100 viral copies (lower right). (D) Replicates of the titrated viral copes using LAMP, as measured by Quantifluor fluorescence. (E) The sensitivity and specificity of the LAMP assay from 182 patients (118 negative and 62 positive for SARS-CoV-2, as measured by RT-PCR).

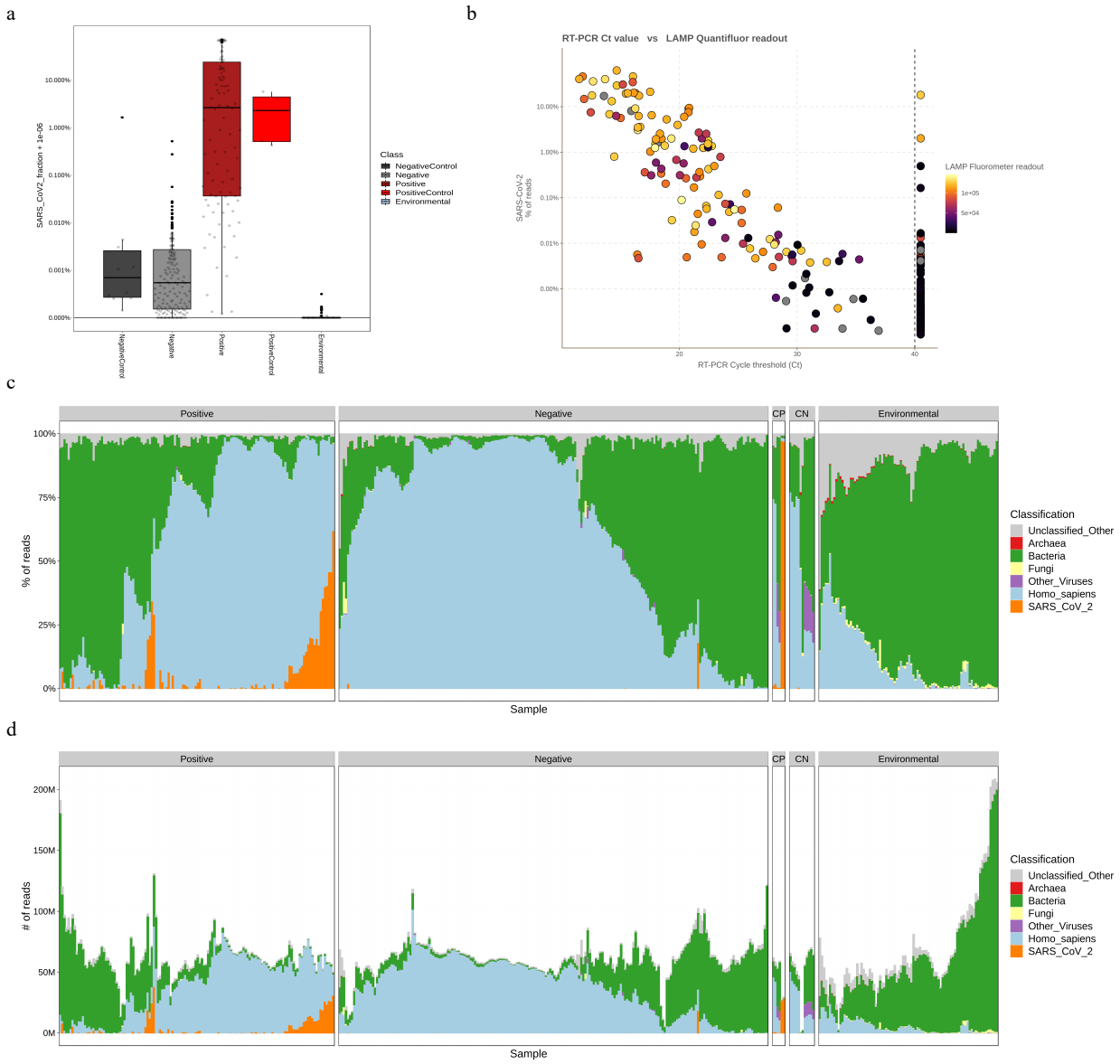


Figure 2. Full transcriptome profiles of SARS-CoV-2 Patients with NGS, RT-PCR, and LAMP. (a) Clinical samples tested by RT-PCR (Positive, dark red or Negative, light grey) were sequenced and run through the LAMP assay. These results were compared to the buffer blanks (Negative Control), dark grey, Synthetic RNAs or Vero 6 cell extracts with SARS-CoV-2 infection (Positive Controls, light red), and Subway Samples (Environmental, blue). Read proportions are shown on the y-axis. (b) SARS-CoV-2 abundance, as measured with NGS and percentage of reads (y-axis) is compared to the Ct Threshold for RT-PCR (x-axis), with lower Ct values representing higher viral abundance, and the LAMP reaction output (Fluorimeter values, black to yellow scale). (c,d) Read mapping to archaea (red), bacteria (blue), fungi (turquiose), chordates (brown), viruses (purple), human (green), and SARS-CoV-2 (tan) across the clinical controls (CN, CP), environmental samples (E), RT-PCR negative (N), and RT-PCR positive (P) samples.

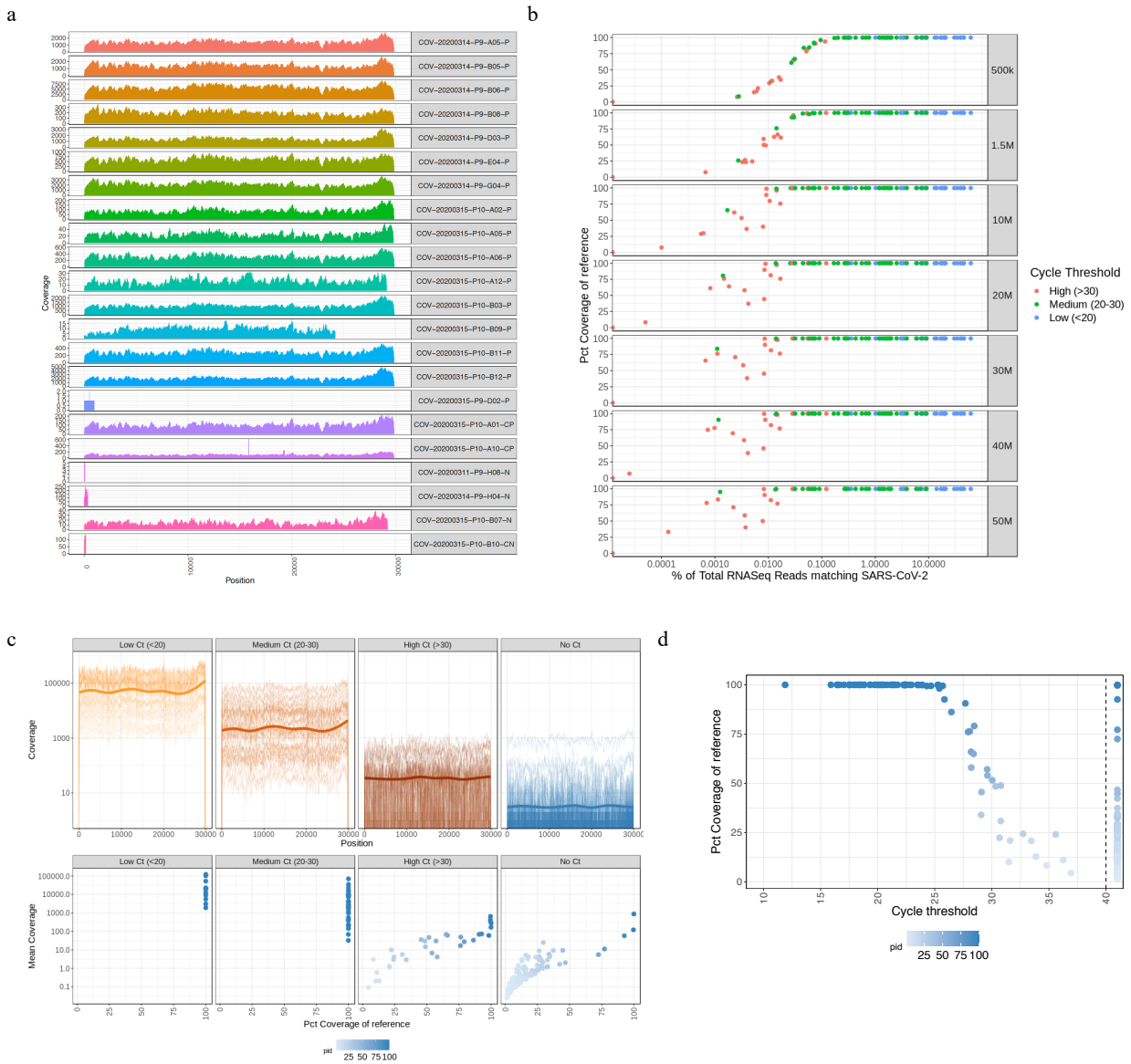


Figure 3: Viral genomes from RNA-seq data and titration of coverage (a) The coverage plot across the SARS-CoV-2 genomes (viral coordinate on bottom, colored by sample) from a representative set of clinical positive samples. CN samples and N samples are clinical negative (buffer), P samples are RT-PCR positive, and CP samples are Vero 6 cells with virus (CP). (b) Downsampling (right annotation) of the samples and mapping to the SARS-CoV-2 genome to gauge the percent coverage (y-axis) as a function of the viral quantification by RT-PCR (Ct thresholds, low <20, medium 20-30, and high>30). (c) Average coverage statistics for the low, medium, and high Ct samples, as well as the mean coverage for each of these samples. (d) The cycle threshold (x-axis) vs. the coverage of the genome (y-axis and color depth) for the total RNA-seq.

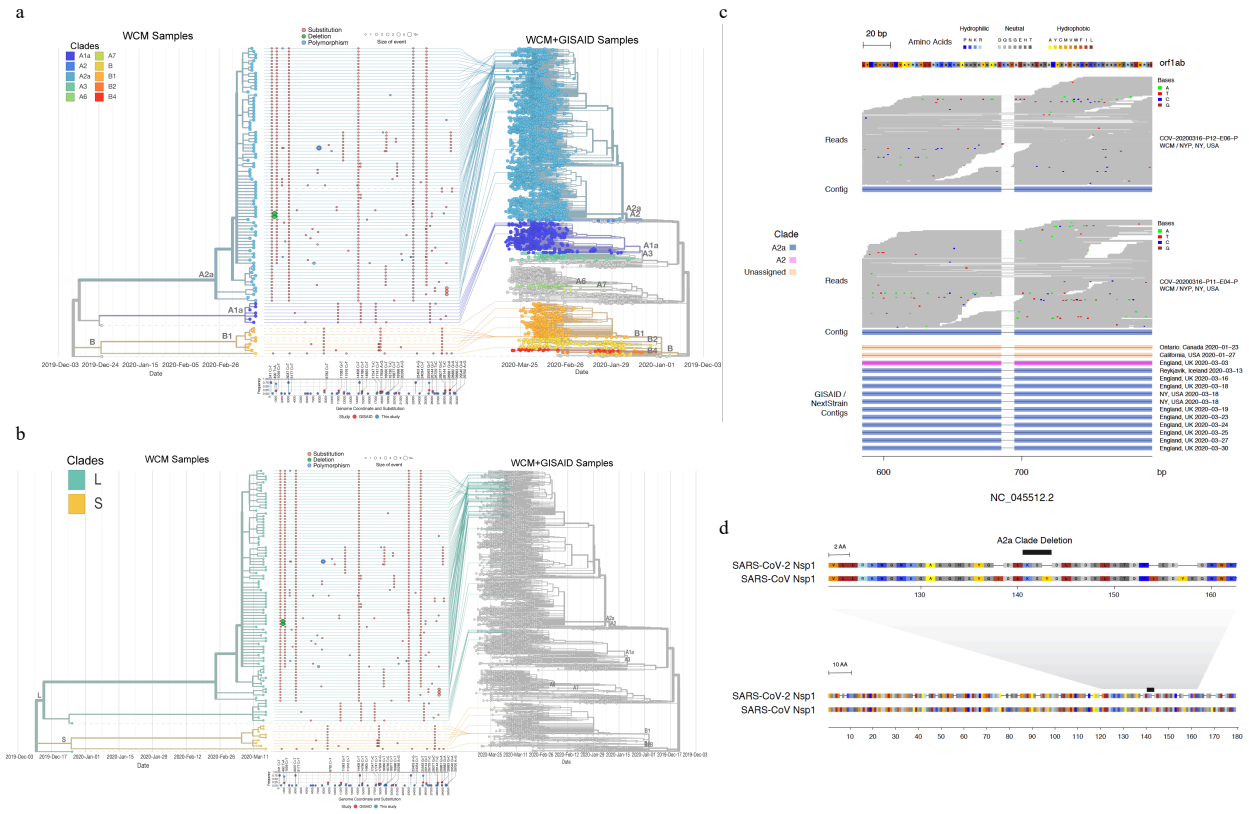


Figure 4: Variation and phylogenetics of NYC COVID-19 Cases. (a) The phylogenetic placement of these COVID samples is shown on the tree (left) and the global map of known COVID genomes (right). Genetic variants called from the RNA-seq data a range of variants that are distinct from the Wuhan reference strain, and the samples from this study, highlighted in blue, show enrichment for European and Asian alleles. (b) Proportion of the L (green) and S strain (yellow) are shown for the NYC viruses. Phylogeny of samples from this study on the left and total GISAID samples on the right, with a map of variants in this study's samples in the middle, colored by event type and sized by number of nucleotides impacted. Annotation track on the bottom shows frequency of alternate alleles in this study and in the GISAID database. (c) Two patients that showed the 9-bp deletion in ORRF1b (NSP1) were confirmed in the GISAID database (below). (d) The ORF1b deletion represents a clade-specific (A2a) deletion that is in a region of the CoV-2 genome unique relative to CoV-1.

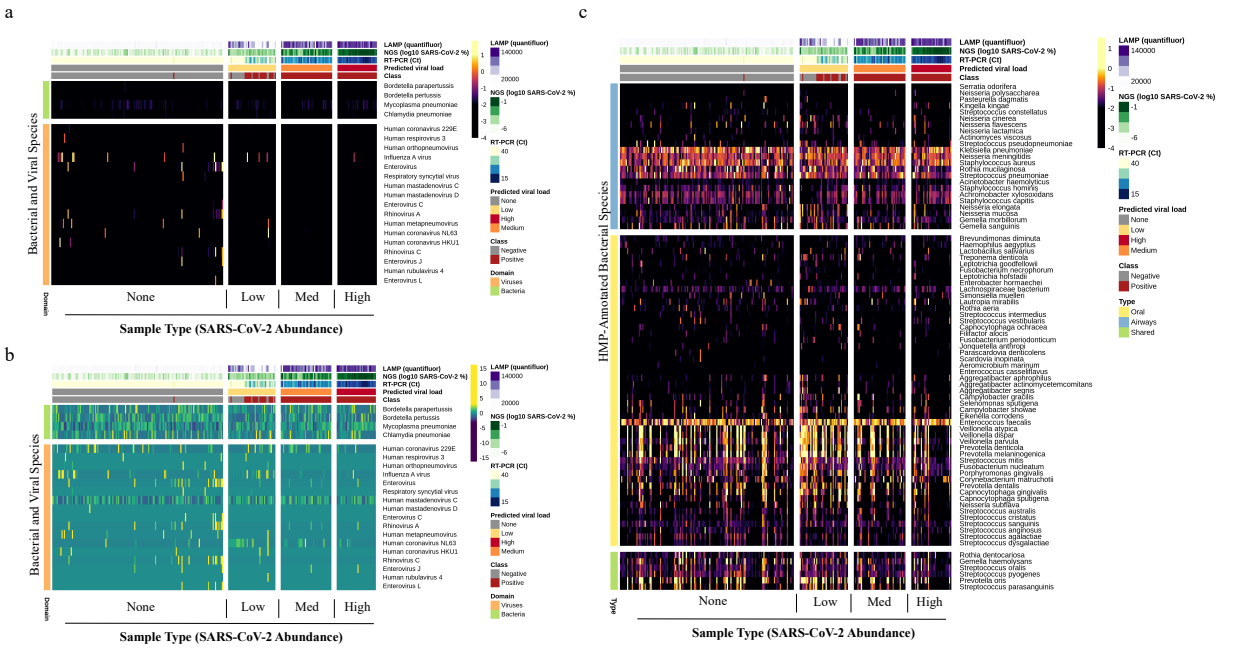


Figure 5: Metatranscriptome profiles of the patient cohorts. (top row, all panels) Samples were quantified by a range of viral detection methods, including LAMP (Quantifluor), RNA-seq (log10 SARS-CoV-2 % of reads), and RT-PCR (Ct values) to create a three-tier range of viral load for the positive samples (right) compared to the clinically-annotated negative samples (class, red or grey). (bottom) The detected microbial species (horizontal lines) are plotted as a z-score of their divergence from the patient cohort (vertical lines). (a) Common respiratory pathogens plotted as a $\log_{10}$ abundance of mapped reads, with each organism as a line and each vertical column as a patient. (b) The same plot as (a), but with z-score for statistical divergence. (c) The same layout as samples, but with bacteria from annotated from the Human Microbiome Project (HMP) as normal airway (blue, top portion), oral (yellow, middle set), or both oral and airway flora (green, bottom).

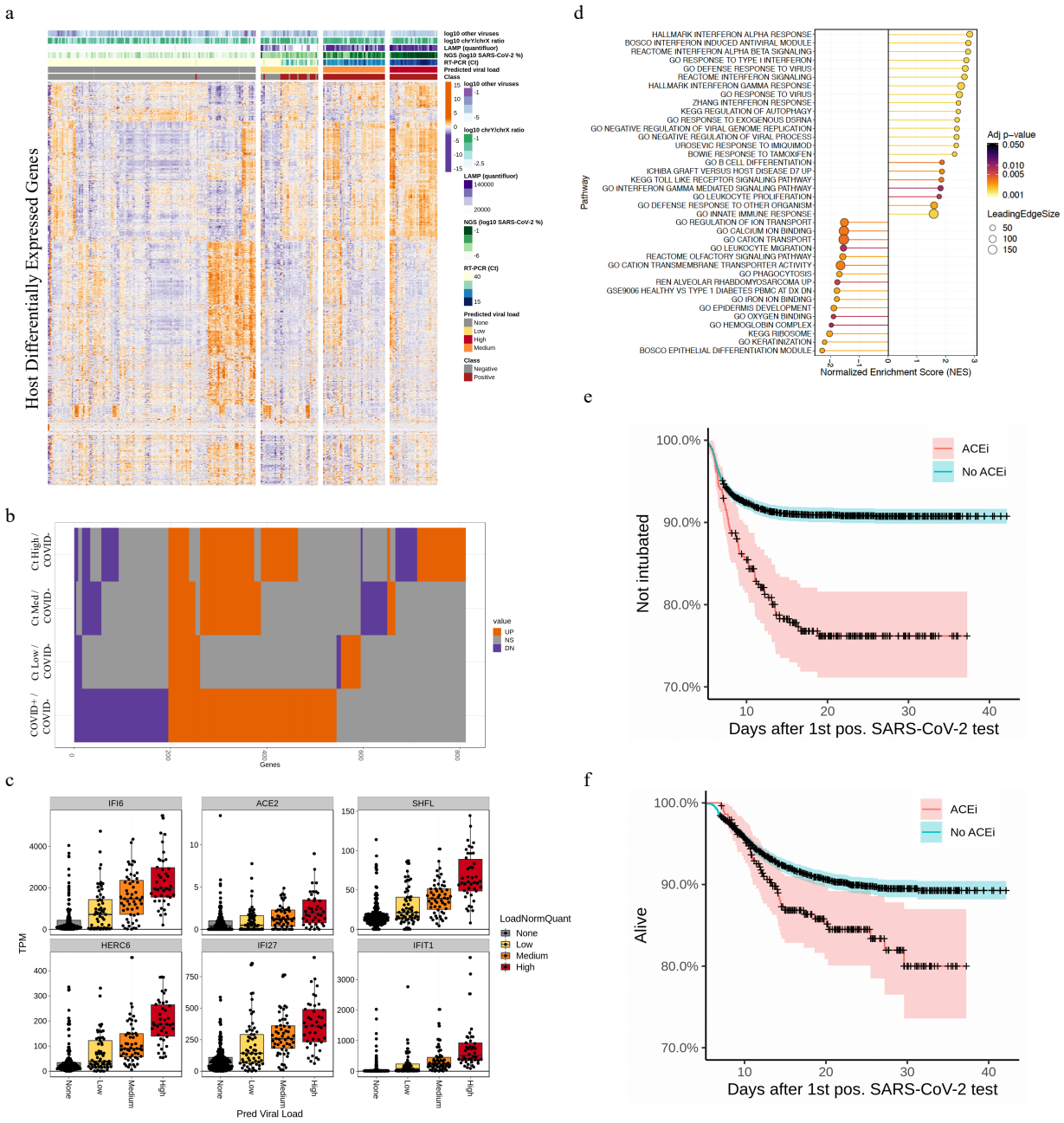


Figure 6: Host transcriptome responses and risk to SARS-CoV-2. (top row) Samples were quantified by a range of viral detection methods, including LAMP (Quantifluor), RNA-seq (log10 SARS-CoV-2 % of reads), and RT-PCR (Ct values) to create a three-tier range of viral load for the positive samples (right) compared to the clinically-annotated negative samples (class, red or grey). (bottom) The differentially expressed genes of COVID+ patients compared to COVID- patients showed up-regulated (orange) genes as well as down-regulated (purple) genes. (b) Up-regulated genes, with boxplots across all samples, include IFI6, ACE2, SHFL, HERC6, IFI27, and IFIT1, based on data from (c), which is the total set of DEGs. The full set is shown in an intersecting heat map, with a core set of up-regulated genes (orange) distinct form the set of down-regulated genes (purple), compared to genes that are not significantly differently

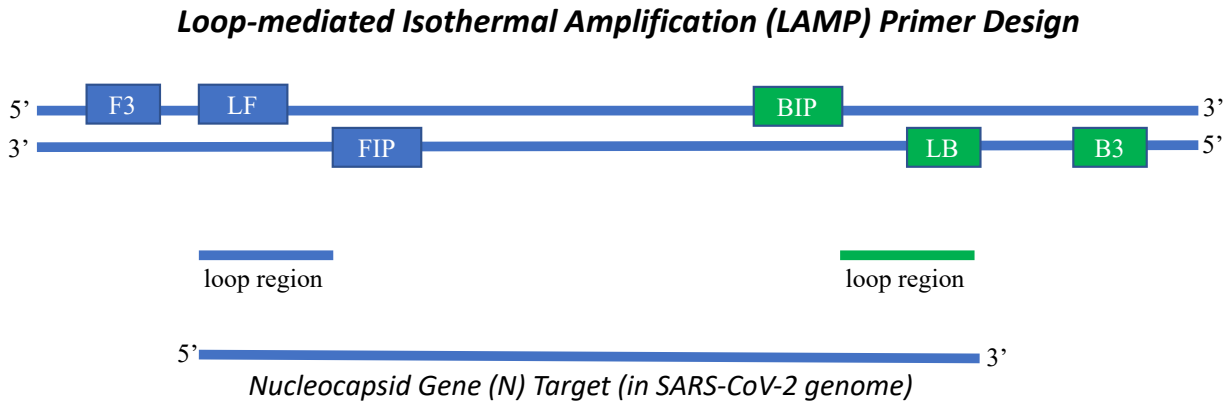
expressed (grey) in any comparison (DESeq2, q-value <0.01, |logFC| >0.58). (d) Significantly different gene ontology (GO) pathways between COVID-POS and COVID-NEG patients include interferon response and host response to infection, as well as some ion and heme transport mechanisms. (e)Survival curves for requiring mechanical respiration (identified by intubation procedure notes) (e) and mortality (f). Patients with a history of ACE inhibitor exposure were more likely to require intubation (HR=2.63 95%CI 2.01-3.43, p=1.22E-12; (e) and less likely to survive (HR=1.68 95%CI: 1.22-2.31, p=1.42E-03; (f). Because several individuals were intubated shortly before they first tested positive for SARS-CoV-2 infection, each first positive test was set back by seven days to account for testing delays.

Table 1. Baseline characteristics of SARS-CoV-2 suspected cohort. General population refers to a comparison cohort of individuals administered drugs at NYP/CUIMC in 2019 who were not later tested for SARS-COV-2 infection.

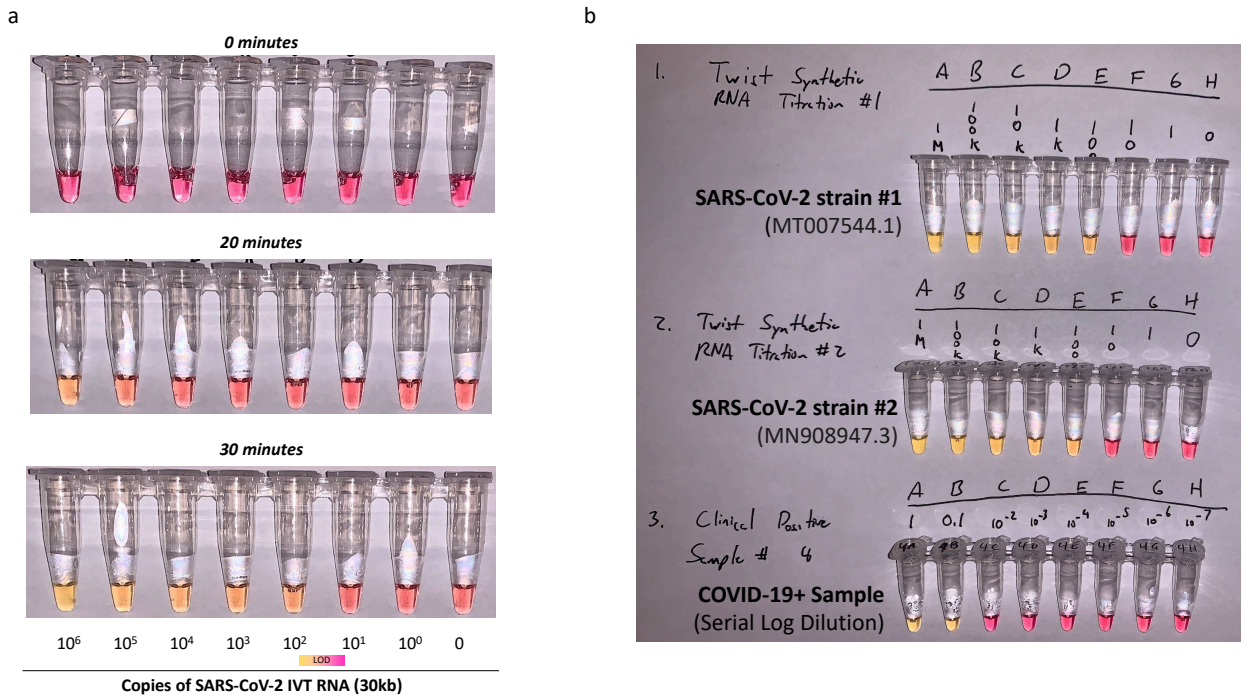
	General Population	**Tested**	**COV+**	**COV+/Intubated**	**COV+/Died**
N (% of tested)	279487	8278 (100%)	4574 (55.3%)	493 (6%)	423 (5.1%)
ACEi	20017 (7.2%)	370 (4.5%)	290 (6.3%)	69 (14%)	42 (9.9%)
Median age (95%)	28.8 (1.4-49.3)	55.7 (0.5-91.3)	62.1 (23.3-92.5)	65.1 (19.3-89.3)	79.2 (49-96.1)
Male	118555 (42.4%)	3806 (46%)	2413 (52.8%)	308 (62.5%)	251 (59.3%)
Black/ African-American	80449 (28.8%)	917 (11.1%)	531 (11.6%)	71 (14.4%)	59 (13.9%)
Caucasian	24267 (8.7%)	1392 (16.8%)	682 (14.9%)	89 (18.1%)	89 (21%)
Asian or Pacific Islander	9 (0.003%)	88 (1.1%)	30 (0.7%)	3 (0.6%)	2 (0.5%)
Other race	167 (0.06%)	1934 (23.4%)	1170 (25.6%)	165 (33.5%)	164 (38.8%)
Missing race	174595 (62.5%)	3947 (47.7%)	2161 (47.2%)	165 (33.5%)	109 (25.8%)
Hispanic/Latino	36905 (13.2%)	1750 (21.1%)	1088 (23.8%)	171 (34.7%)	159 (37.6%)
Non-Hispanic/Latino	87562 (31.3%)	1547 (18.7%)	746 (16.3%)	85 (17.2%)	87 (20.6%)
Other listed ethnicity	155020 (55.5%)	1034 (12.5%)	579 (12.7%)	72 (14.6%)	68 (16.1%)
Hypertension	104925 (38.4%)	2168 (26.2%)	1368 (29.9%)	221 (44.8%)	260 (61.5%)
CAD/CHD	155561 (56.9%)	816 (9.9%)	503 (11%)	84 (17%)	115 (27.2%)

Diabetes	46908 (17.2%)	1274 (15.4%)	834 (18.2%)	151 (30.6%)	174 (41.1%)
Overweight	39192 (14.3%)	452 (5.5%)	259 (5.7%)	43 (8.7%)	38 (9%)
No risk factors	110460 (40.4%)	5830 (70.4%)	3052 (66.7%)	248 (50.3%)	150 (35.5%)
Median IL-6 (50%)	-	30.6 (8-94.1)	31.1 (8-94.9)	68.2 (16-178)	68 (18-153.8)

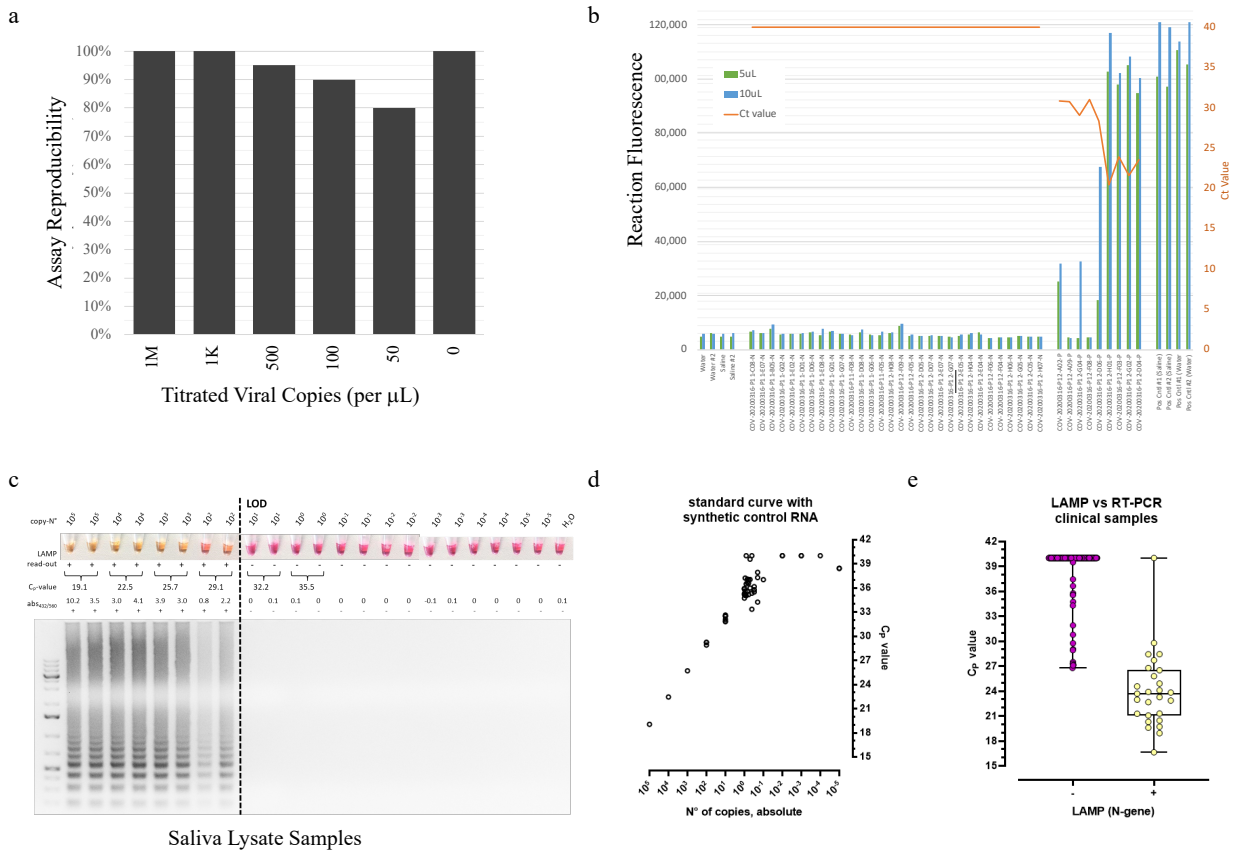
Supplemental Figures



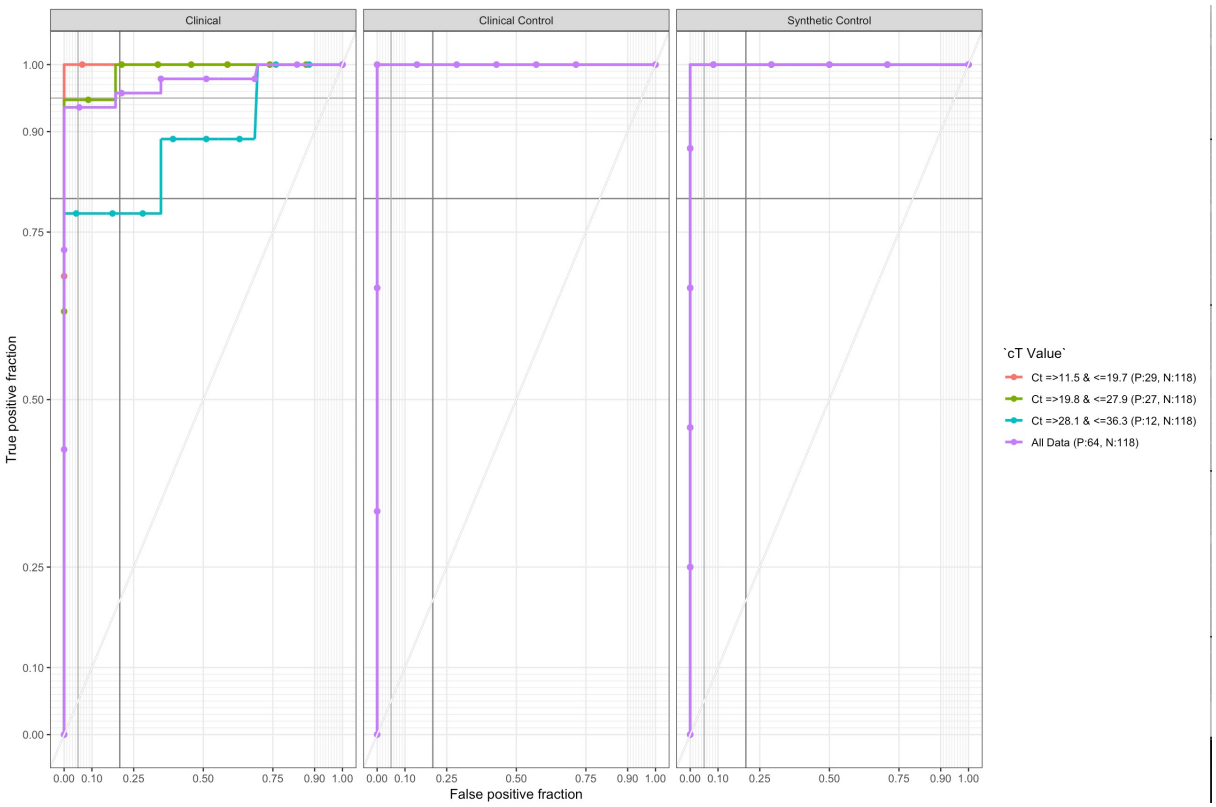
Supplemental Figure 1: LAMP Primer Design. PrimerDesigner (v4) was used to create a set of six primers that would specifically target the nucleocapsid gene (N) in the COVID-19 genome. Primer sequences are listed in the methods section.



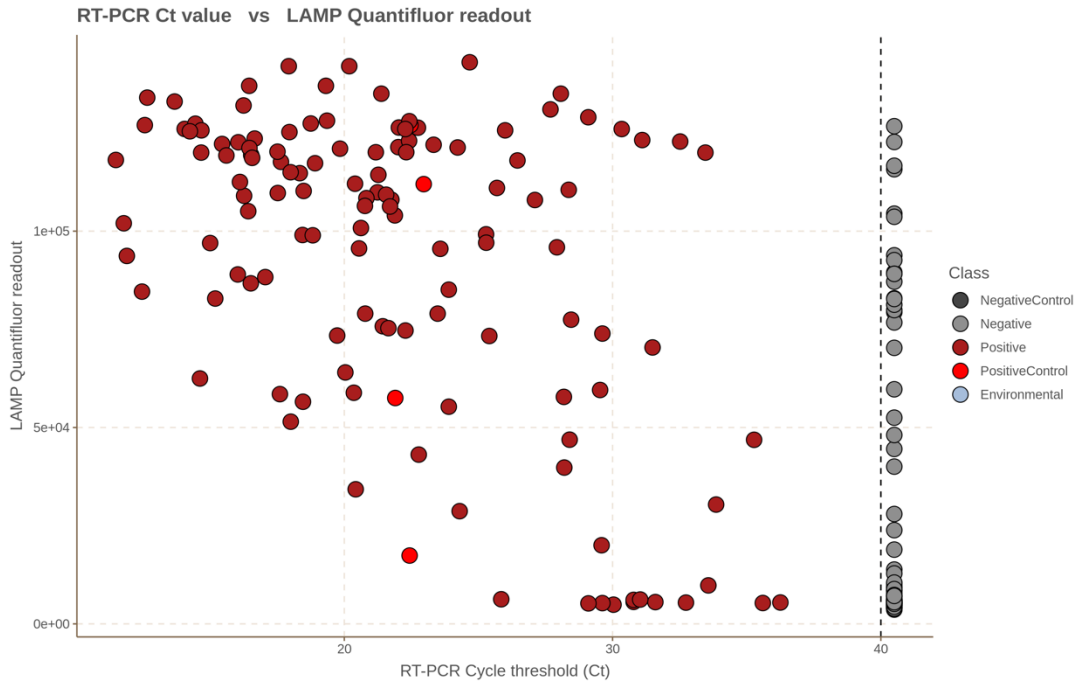
Supplemental Figure 2. Additional Testing and Titration of the LAMP Assay with Synthetic and Clinical Samples. Samples were prepared using the LAMP protocol with a reaction time of 30 minutes. Reaction progress was measured (a) from 0, 20, and 30 minutes. (b) This was repeated for both Twist COVID-19 synthetic RNAs (*MT007544.1*, top and MN908947.3, middle*)* from 1 million molecules of virus (10^6), then titrated down by log10 dilutions. Limit of Detection (LOD) range is shown with a gradient after 30 minutes between 10 and 100 viral copies. (bottom) A clinically positive sample that was not detectible by Qubit (<0.05 ng/mL) was nonetheless detected by LAMP, in accordance with detection of low viral titer samples.



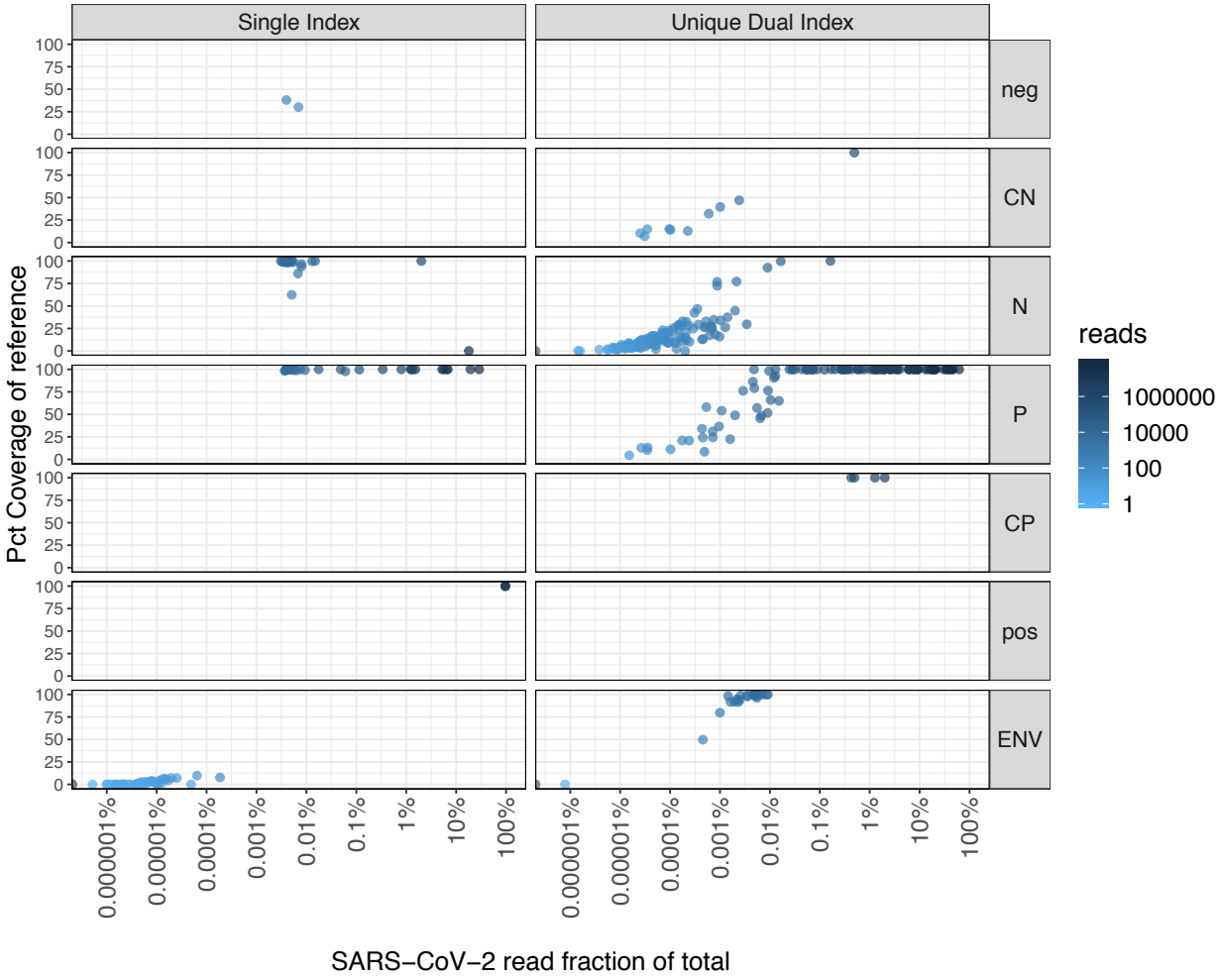
Supplemental Figure 3. Reproducibility, sensitivity, and specificity for the LAMP Assay. (a) Testing with a synthetic SARS-CoV-2 RNA that was serially titrated and measured in replicates (n=10) showed 100% and 95% reproducibility at 1,000 and 500 copies, respectively, with lower rates at lower viral titers. (b) Replicates of a clinically positive (by RT-PCR) sample at 10 uL (blue) compared to 5uL (green) showed high concordance, with greater sensitivity with increased reaction volume. (c) Whole oropharyngeal swab lysates from clinical positive (Cp-value >0) and negative samples (Cp = NA) were used to test the LAMP reaction. (d) The standard curve with synthetic RNA was also tested relative to absolute number of copies (x-axis) and the Cp value (y-axis). (e) The Cp value for the LAMP positive (+, right) and negative (-, left) were compared to the Cp value from RT-PCR (y-axis).



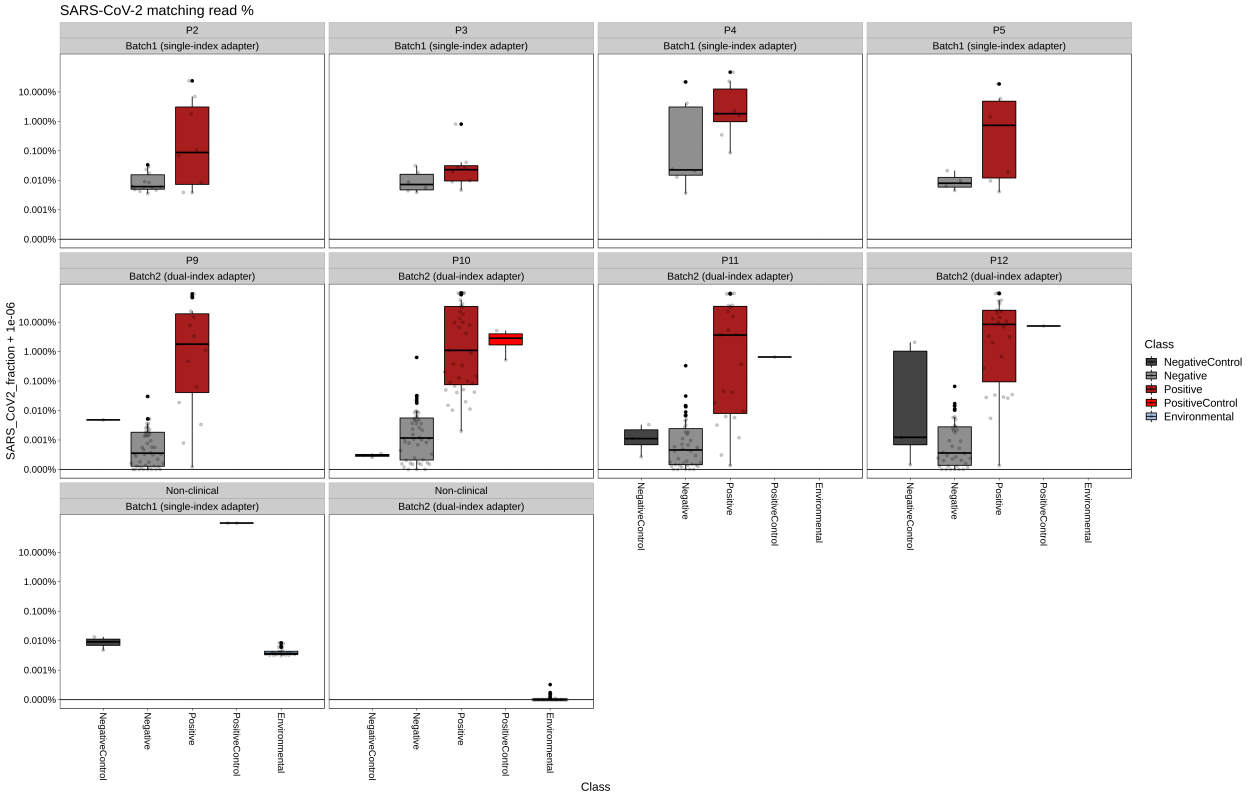
Supplemental Figure 4. Ct Thresholds for the LAMP Assay. With increasing viral load, as measured by RT-PCR, the LAMP assay shows an increased sensitivity (y-axis) and specificity (y-axis). For high and medium viral count samples (Ct < 28, red and green line), we see 100% sensitivity and specificity, and for low viral count, we see nearly 80% sensitivity and 100% specificity (blue). Across all data (purple), we see 94% sensitivity and 100% specificity.



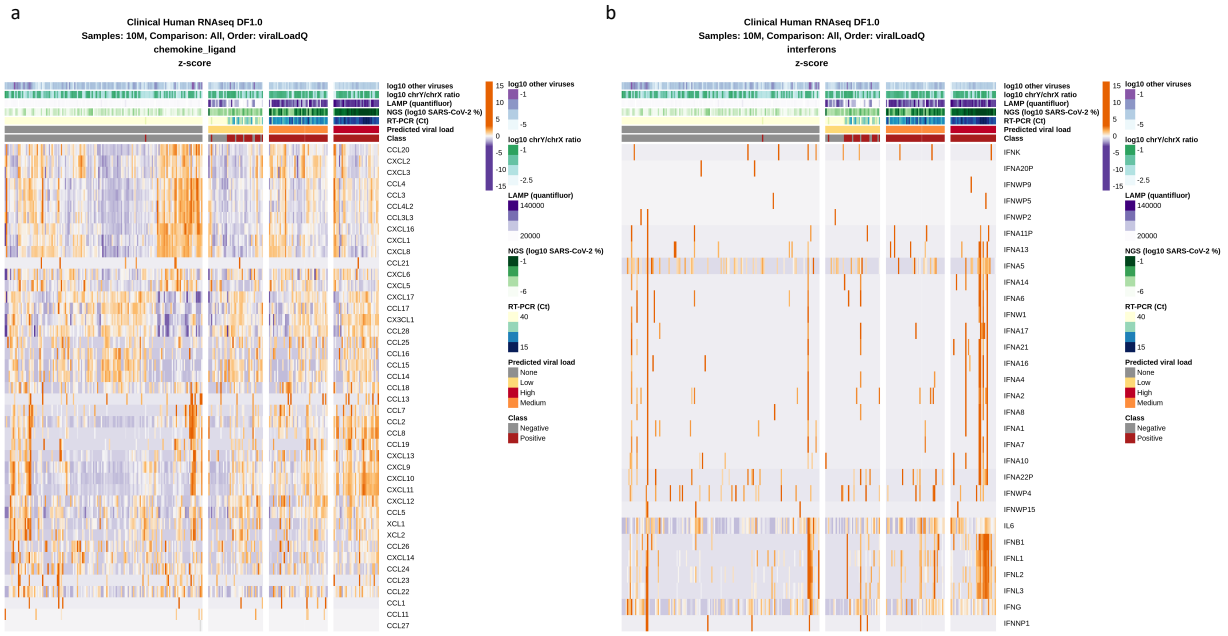
Supplementary Figure 5. Correlation between LAMP Reaction Output and RT-PCR. Clinical samples tested by RT-PCR (Positive, dark red or Negative, light grey) were run with the LAMP assay and compared to the buffer blanks (Negative Control), dark grey, Synthetic RNAs or Vero 6 cell extracts with SARS-CoV-2 infection (Positive Controls, light red), and Subway Samples (Environmental, blue, lower right). The DNA abundance, as measured with the GloMax Quantifluor (y-axis) is compared to the Ct Threshold for RT-PCR (x-axis), with lower Ct values representing higher viral abundance.



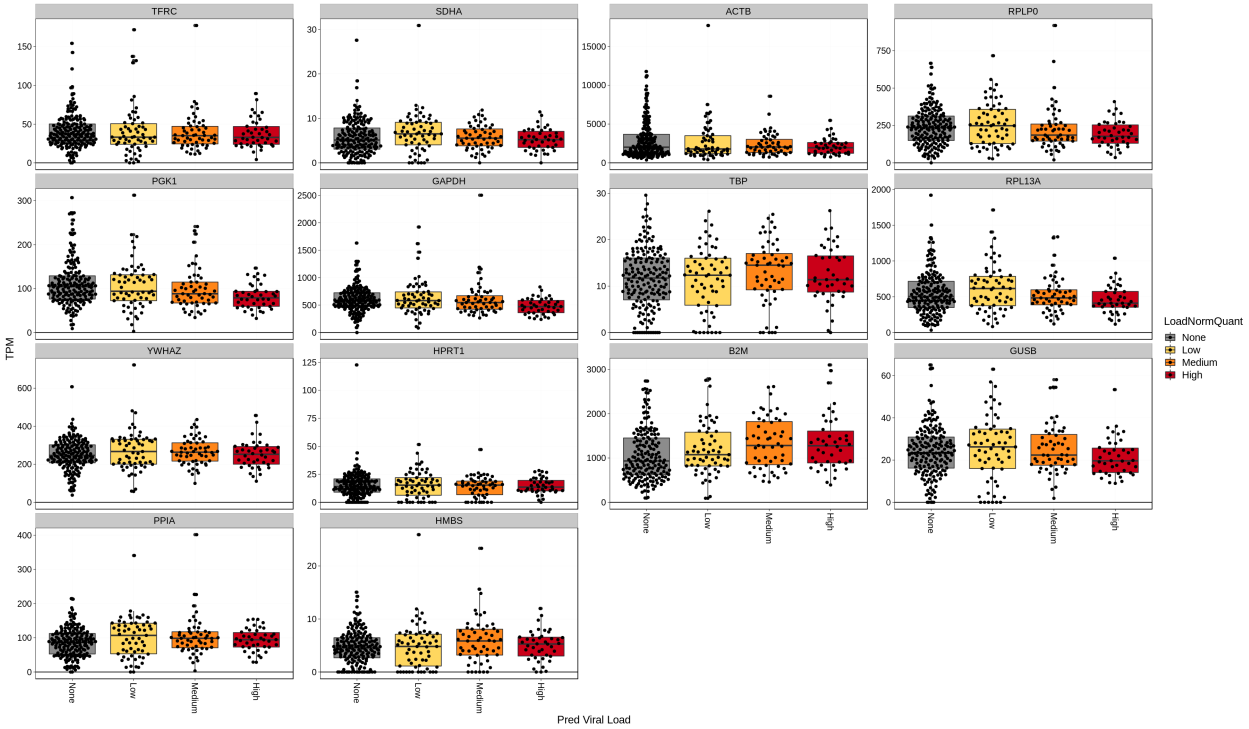
Supplementary Figure 6. Coverage of COVID samples by Index and Sample Type . Clinical samples tested by RT-PCR (Positive, P, Negative, N) were compared to the buffer blanks (Negative Control, N and neg), Synthetic RNAs or Vero 6 cell extracts with SARS-CoV-2 infection (CP and pos) and Subway Samples (Environmental, ENV). The %identity to the SAS-CoV-2 reference (y-axis) is shown relative to the proportion of reads that mapped from NGS (x-axis). The unique, dual-index runs (left panels) are shown relative to the single-index runs (right panels).



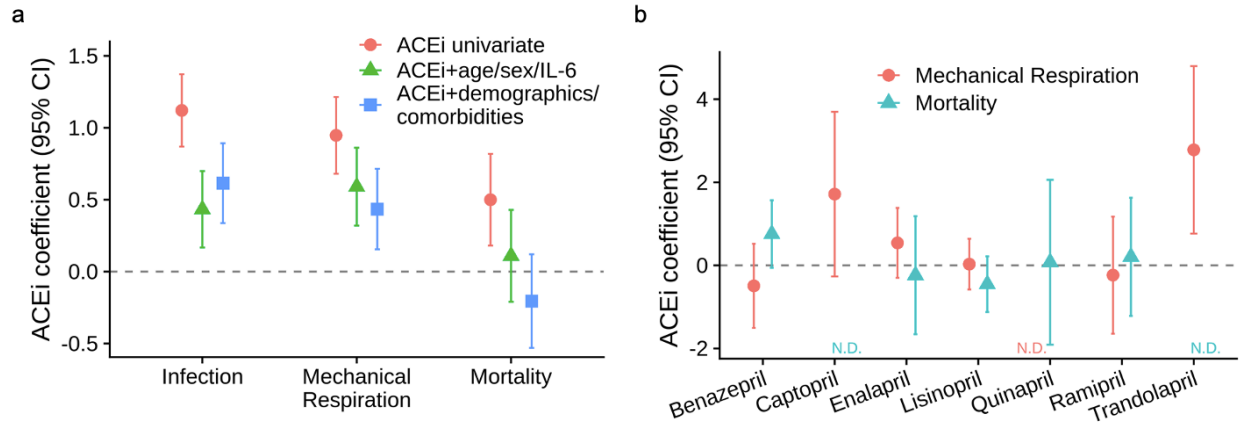
Supplemental Figure 7. Proportion of RNA-seq reads mapping SARS-CoV-2. Clinical samples tested by RT-PCR (Positive, dark red or Negative, light grey) were sequenced and compared to the buffer blanks (Negative Control), dark grey, Synthetic RNAs or Vero 6 cell extracts with SARS-CoV-2 infection (Positive Controls, light red), and Subway Samples (Environmental, blue). Read totals are shown on the y-axis. Differences between single index barcodes are plotted across each of the plates of samples that were processed.



Supplemental Figure 8. Cytokine and interferon profiles of the host transcriptome. Top rows: samples were quantified by a range of viral detection methods, including LAMP (Quantifluor), RNA-seq (log10 SARS-CoV-2 % of reads), and RT-PCR (Ct values) to create a three-tier range of viral load for the positive samples (right) compared to the clinically-annotated negative samples (class, red or grey). (bottom) The differentially expressed genes of COVID+ patients compared to COVID- patients showed up-regulated (orange) genes as well as down-regulated (purple) genes. (a) Chemokine profiles for the samples (x-axis) is plotted for each related gene (y-axis), and the same plotting function shows in (b) the interform and interleukin gene profiles of IL6.



Supplemental Figure 9. Housekeeping host genes relative to SARS-CoV-2 infection. Clinical samples were sequenced with RNA-seq and quantified to a set of genes for their expression levels. Samples with no virus (grey) were compared to those with low (yellow), medium (orange), and high (red) expression levels, based on RT-PCR.



Supplemental Figure 10. ACEI Multivariate and Comparative Analyses. (a) Regression coefficients for variables indicating exposure/history of exposure to ACE inhibitors for each of the three cohort comparisons: (left) test outcome in a cohort of patients suspected of SARS-CoV-2 infection, (middle) requirement of mechanical respiration in patients who tested positive, (right) mortality in patients who tested positive. Univariate analyses are shown as red circles. The green triangles coefficients are when correcting for age, sex, and baseline IL-6 levels take upon admission. The blue squares are from a model that includes age, sex, and IL-6 as well as comorbidities including CAD/CHD, diabetes, obesity, and self-reported race and ethnicity. (b) Comparison of the effects of different ACE inhibitors. We directly compared the effects of specific ACE inhibitors among those patients with evidence of exposure to ACE inhibitors. We found two significant associations: benazepril was significantly associated with

higher mortality (HR=2.37 95%CI 1.05-5.35, p=3.70E-02) and trandolapril was associated with requirement of mechanical ventilation (HR=15.85 95% CI: 2.11-119.08, p=7.25E-03). N.D. indicates that a result was not displayed due to low sample size and, therefore, very large errors.

Supplementary Table 1: Metatranscriptome Profiles of All Samples. Appended.

Supplementary Table 2: Taxonomic Mis-assignment Filter. The entire SARS-CoV-2 genome was fragmented (wgsim) and then re-mapping to the entire Kraken2 database. These species were then used to flag putative false positives from SARS-CoV-2 genome segments that are similar to other organisms.

Supplementary Table 2 : Other Taxonomic Alignments for SARS-CoV-2 (High-risk false positives)

Taxonomic group (domain, phylum, class, order, family, genus, species, strain)	TaxonName	% match
d__Viruses|o__Nidovirales|f__Coronaviridae|g__Betacoronavirus|s__Severe acute respiratory syndrome-related coronavirus	Severe acute respiratory syndrome-related coronavirus (SARS)	95.0813%
d__Viruses|o__Nidovirales|f__Coronaviridae|s__Bat coronavirus BM48-31/BGR/2008	Bat coronavirus BM48-31/BGR/2008	0.3937%
d__Eukaryota|k__Metazoa|p__Chordata|c__Mammalia|o__Primates|f__Hominidae|g__Homo|s__Homo sapiens	Homo sapiens	0.0170%
d__Viruses|o__Nidovirales|f__Coronaviridae|g__Alphacoronavirus|s__Rhinolophus ferrumequinum alphacoronavirus HuB-2013	Rhinolophus ferrumequinum alphacoronavirus HuB-2013	0.0164%
d__Viruses|o__Nidovirales|f__Coronaviridae|g__Betacoronavirus|s__Betacoronavirus 1	Betacoronavirus 1	0.0099%
d__Viruses|o__Nidovirales|f__Coronaviridae|g__Betacoronavirus|s__Murine coronavirus	Murine coronavirus	0.0089%
d__Viruses|o__Nidovirales|f__Coronaviridae|g__Betacoronavirus|s__Rousettus bat coronavirus GCCDC1	Rousettus bat coronavirus GCCDC1	0.0059%
d__Viruses|o__Nidovirales|f__Coronaviridae|s__Bat coronavirus	Bat coronavirus	0.0049%
d__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium botulinum	Clostridium botulinum	0.0045%
d__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium perfringens	Clostridium perfringens	0.0037%
d__Bacteria|p__Fusobacteria|c__Fusobacteriia|o__Fusobacteriales|f__Fusobacteriaceae|g__Fusobacterium|s__Fusobacterium varium	Fusobacterium varium	0.0027%
d__Viruses|o__Nidovirales|f__Coronaviridae|g__Alphacoronavirus|s__Myotis ricketti alphacoronavirus Sax-2011	Myotis ricketti alphacoronavirus Sax-2011	0.0025%
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacterales|f__Erwiniaceae|g__Buchnera|s__Buchnera aphidicola	Buchnera aphidicola	0.0024%
d__Bacteria|p__Firmicutes|c__Bacilli|o__Bacillales|f__Thermoactinomycetaceae|g__Thermoactinomyces|s__Thermoactinomyces vulgaris	Thermoactinomyces vulgaris	0.0023%
d__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Streptococcus|s__Streptococcus pyogenes	Streptococcus pyogenes	0.0023%
d__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Rickettsiales|f__Anaplasmataceae|g__Ehrlichia|s__Ehrlichia sp. HF	Ehrlichia sp. HF	0.0018%
d__Bacteria|p__Firmicutes|c__Bacilli|o__Bacillales|f__Bacillaceae|g__Bacillus|s__Bacillus cereus	Bacillus cereus	0.0017%
d__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|g__Candidatus Azobacteroides|s__Candidatus Azobacteroides pseudotrichonymphae	Candidatus Azobacteroides pseudotrichonymphae	0.0017%
d__Viruses|o__Nidovirales|f__Coronaviridae|g__Betacoronavirus|s__Rousettus bat coronavirus HKU9	Rousettus bat coronavirus HKU9	0.0015%
d__Viruses|s__Beihai picorna-like virus 116	Beihai picorna-like virus 116	0.0015%
d__Bacteria|p__Proteobacteria|c__Epsilonproteobacteria|o__Campylobacterales|f__Helicobacteraceae|g__Helicobacter|s__Helicobacter pylori	Helicobacter pylori	0.0014%
d__Bacteria|p__Proteobacteria|c__Betaproteobacteria|o__Burkholderiales|f__Oxalobacteraceae|g__Janthinobacterium|s__Janthinobacterium sp. B9-8	Janthinobacterium sp. B9-8	0.0014%
d__Eukaryota|p__Apicomplexa|c__Aconoidasida|o__Haemosporida|f__Plasmodiidae|g__Plasmodium|s__Plasmodium coatneyi	Plasmodium coatneyi	0.0014%
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacterales|f__Yersiniaceae|g__Serratia|s__Serratia symbiotica	Serratia symbiotica	0.0013%
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter baumannii	Acinetobacter baumannii	0.0013%
d__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium septicum	Clostridium septicum	0.0013%
d__Viruses|o__Nidovirales|f__Coronaviridae|g__Betacoronavirus|s__Tylonycteris bat coronavirus HKU4	Tylonycteris bat coronavirus HKU4	0.0012%
d__Bacteria|p__Firmicutes|c__Bacilli|o__Bacillales|f__Staphylococcaceae|g__Staphylococcus|s__Staphylococcus cohnii	Staphylococcus cohnii	0.0012%
d__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Caloranaerobacter|s__Caloranaerobacter azorensis	Caloranaerobacter azorensis	0.0012%
d__Bacteria|p__Cyanobacteria|o__Nostocales|f__Calotrichaceae|g__Calothrix|s__Calothrix sp. NIES-2098	Calothrix sp. NIES-2098	0.0012%
d__Viruses|o__Nidovirales|f__Coronaviridae|g__Betacoronavirus|s__Bat Hp-betacoronavirus Zhejiang2013	Bat Hp-betacoronavirus Zhejiang2013	0.0011%
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacterales|f__Erwiniaceae|g__Erwinia|s__Candidatus Erwinia haradaeae	Candidatus Erwinia haradaeae	0.0011%
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Thiotrichales|f__Thiotrichaceae|g__Thioploca|s__Thioploca ingrica	Thioploca ingrica	0.0011%
d__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Rhodospirillales|f__Acetobacteraceae|g__Bombella|s__Bombella sp. KACC 21507	Bombella sp. KACC 21507	0.0011%
d__Bacteria|p__Cyanobacteria|o__Chroococcales|f__Aphanothecaceae|g__Candidatus Atelocyanobacterium|s__Candidatus Atelocyanobacterium thalassa	Candidatus Atelocyanobacterium thalassa	0.0011%
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacterales|f__Erwiniaceae|g__Wigglesworthia|s__Wigglesworthia glossinidia	Wigglesworthia glossinidia	0.0010%
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacterales|f__Enterobacteriaceae|g__Enterobacter|s__Enterobacter hormaechei	Enterobacter hormaechei	0.0010%
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Alteromonadales|f__Colwelliaceae|g__Colwellia|s__Colwellia sp. PAMC 20917	Colwellia sp. PAMC 20917	0.0010%
d__Bacteria|p__Proteobacteria|c__Betaproteobacteria|o__Neisseriales|f__Neisseriaceae|g__Neisseria|s__Neisseria flavescens	Neisseria flavescens	0.0010%
d__Bacteria|p__Bacteroidetes|c__Flavobacteriia|o__Flavobacteriales|f__Flavobacteriaceae|g__Lutibacter|s__Lutibacter profundi	Lutibacter profundi	0.0010%

Supplementary Table 3: Differentially Expressed Genes in COVID-19+/- patients. Appended.

Supplemental Table 4: Gene Ontology Pathways.
Appended

Supplementary Table 5: COVID-19 Rapid Colorimetric LAMP Detection Test: N Primers Specificity

Supplemental Table 5: COVID-19 Rapid Colorimetric LAMP Detection Test: N Primers						
Organism	Range % Homology with F3 Primer	Range % Homology with B3 Primer	Range % Homology with FIP Primer	Range % Homology with BIP Primer	Range % Homology with LF Primer	Range % Homology with LB Primer
Human coronavirus 229E (https://www.ncbi.nlm.nih.gov/nuccore/NC_002645.1)	0%	0%	0%	50%	0%	0%
Human coronavirus OC43 (https://www.ncbi.nlm.nih.gov/assembly/GCF_003972325	0%	0%	0%	0%	0%	0%
Human coronavirus HKU1 (https://www.ncbi.nlm.nih.gov/nuccore/NC_006577.2)	0%	0%	0%	0%	0%	0%
Human coronavirus NL63 (https://www.ncbi.nlm.nih.gov/nuccore/NC_005831.2)	0%	0%	0%	0%	0%	0%
SARS-coronavirus (https://www.ncbi.nlm.nih.gov/nuccore/NC_004718.3)	100%	85%	46%	43-50%	88%	100%
MERS-coronavirus (https://www.ncbi.nlm.nih.gov/nuccore/NC_038294.1)	0%	0%	0%	0%	0%	0%
SARS-coronavirus 2 (https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2)	100%	100%	46%-54%	45%-55%	100%	100%
Bordetella pertussis 18323, complete sequence	0%	0%	37%-41%	65%-73%	72%	0%
Candida albicans SC5314 chromosome 1 sequence	0%	65%-75%	0%	35%	0%	0%
Candida albicans SC5314 chromosome 2 sequence	0%	0%	0%	38%	64%	86%
Candida albicans SC5314 chromosome 3 sequence	0%	0%	0%	0%	64%	76%
Candida albicans SC5314 chromosome 4 sequence	0%	0%	0%	0%	68%	0%
Candida albicans SC5314 chromosome 5 sequence	0%	65%	0%	0%	0%	0%
Candida albicans SC5314 chromosome 7 sequence	0%	65%	0%	35%	0%	0%
Candida albicans SC5314 chromosome R sequence	0%	0%	0%	0%	68%	0%
Chlamydia pneumoniae TW-183, complete sequence	79%	0%	0%	0%	0%	0%
Haemophilus influenzae genome assembly NCTC8143, chromosome : 1	0%	0%	0%	0%	64%-72%	0%
Legionella pneumophila subsp. pascullei strain NCTC12273, chromosome: 1	0%	65%-75%	0%	0%	72%	0%
Mycobacterium tuberculosis H37Rv, complete genome	0%	0%	37%-59%	0%	0%	0%
Mycoplasma pneumoniae FH chromosome, complete genome	0%	0%	0%	0%	72%	0%
Pneumocystis jirovecii RU7 chromosome Unknown supercont1.14	0%	0%	0%	0%	76%	0%
Pneumocystis jirovecii RU7 chromosome Unknown supercont1.15	0%	0%	0%	0%	64%-72%	0%
Pneumocystis jirovecii RU7 chromosome Unknown supercont1.8	0%	85%	0%	0%	64%	0%
Pneumocystis jirovecii RU7 chromosome Unknown supercont1.9	0%	0%	0%	0%	64%	0%
Rothia mucilaginosa DY-18 DNA, complete genome	0%	80%	0%	0%	0%	62%
Staphylococcus epidermidis ATCC 12228, complete sequence	0%	75%	0%	35%	76%	0%
Streptococcus pneumoniae genome assembly NCTC7465, chromosome : 1	0%	85%	41%	0%	64%	62%
Streptococcus pyogenes genome assembly NCTC8198, chromosome : 1	0%	0%	0%	35%	64%-68%	62%