Distinguishing Severe Acute Respiratory Syndrome Coronavirus 2 Persistence and Reinfection: A Retrospective Cohort Study

Sarah E. Turbett,[1,2,3,a,] Christopher H. Tomkins-Tinch,[4,5,a] Melis N. Anahtar,[3,4,a] Caitlin M. Dugdale,[1,2,6,a] Emily P. Hyle,[1,2,6]
Erica S. Shenoy,[1,2,7] Bennett Shaw,[1,4,8] Kenechukwu Egbuonu,[9] Kathryn A. Bowman,[1,2,10] Kimon C. Zachary,[1,2,7] Gordon C. Adams,[4]
David C. Hooper,[1,2,7] Edward T. Ryan,[1,2,11] Regina C. LaRocque,[1,2] Ingrid V. Bassett,[1,2,6] Virginia A. Triant,[1,2,12] Katherine J. Siddle,[4]
Eric Rosenberg,[1,2,3] Pardis C. Sabeti,[13] Stephen F. Schaffner,[4,5,11] Bronwyn L. MacInnis,[4,11,b] Jacob E. Lemieux,[1,2,4,b] and Richelle C. Charles[1,2,11,b]

[1]Division of Infectious Diseases, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; [2]Harvard Medical School, Boston, Massachusetts, USA; [3]Department of Pathology, Massachusetts General Hospital (MGH), Boston, Massachusetts, USA; [4]Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, Massachusetts, USA; [5]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA; [6]Medical Practice Evaluation Center, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; [7]Infection Control Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; [8]David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, California, USA; [9]Massachusetts Institute of Technology, Boston, Massachusetts, USA; [10]Ragon Institute of MGH, MIT, and Harvard, Cambridge, Massachusetts, USA; [11]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA; [12]Division of General Internal Medicine, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; and [13]FAS Center for Systems Biology, Harvard University, Boston, Massachusetts, USA

Background. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) reinfection is poorly understood, partly because few studies have systematically applied genomic analysis to distinguish reinfection from persistent RNA detection related to initial infection. We aimed to evaluate the characteristics of SARS-CoV-2 reinfection and persistent RNA detection using independent genomic, clinical, and laboratory assessments.

Methods. All individuals at a large academic medical center who underwent a SARS-CoV-2 nucleic acid amplification test (NAAT) ≥45 days after an initial positive test, with both tests between 14 March and 30 December 2020, were analyzed for potential reinfection. Inclusion criteria required having ≥2 positive NAATs collected ≥45 days apart with a cycle threshold (Ct) value <35 at repeat testing. For each included subject, likelihood of reinfection was assessed by viral genomic analysis of all available specimens with a Ct value <35, structured Ct trajectory criteria, and case-by-case review by infectious diseases physicians.

Results. Among 1569 individuals with repeat SARS-CoV-2 testing ≥45 days after an initial positive NAAT, 65 (4%) met cohort inclusion criteria. Viral genomic analysis characterized mutations present and was successful for 14/65 (22%) subjects. Six subjects had genomically supported reinfection, and 8 subjects had genomically supported persistent RNA detection. Compared to viral genomic analysis, clinical and laboratory assessments correctly distinguished reinfection from persistent RNA detection in 12/14 (86%) subjects but missed 2/6 (33%) genomically supported reinfections.

Conclusions. Despite good overall concordance with viral genomic analysis, clinical and Ct value-based assessments failed to identify 33% of genomically supported reinfections. Scaling-up genomic analysis for clinical use would improve detection of SARS-CoV-2 reinfections.

Keywords. COVID-19; SARS-CoV-2 reinfection; persistent detection.

Nucleic acid amplification tests (NAATs) can detect severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA weeks to months after initial infection [1, 2]. Therefore, positive NAATs obtained several weeks after an initial diagnosis pose a clinical dilemma—these results may reflect delayed clearance of non-viable virus, chronic active infection with viable virus, or SARS-CoV-2 reinfection, each of which carries different implications for treatment, contact tracing, and transmission-based precautions [3–6]. In addition to aiding clinical decision making, distinguishing reinfection from persistent RNA detection related to the initial infection is of public health importance, because reinfections may indicate viral escape due to waning immunity or signal the emergence of novel variants [5, 7]. However, without viral genomic analysis, which is not yet widely available for clinical use, it is challenging to distinguish SARS-CoV-2 reinfection from persistent RNA detection using routinely available clinical and laboratory data alone [3–5].

Previous attempts to identify SARS-CoV-2 reinfection incidence and associated clinical characteristics have mostly relied

on time-based definitions for reinfection (ie, having a repeat positive test at some interval after diagnosis), rather than genomic analysis, leading to potential misclassification of reinfection cases [2, 8–10]. The Centers for Disease Control and Prevention (CDC) investigative criteria for suspected cases of SARS-CoV-2 reinfection prioritizes genomic evaluation of viral specimens from individuals with a repeat positive NAAT ≥45 days after initial diagnosis to distinguish reinfection from persistent RNA detection (Supplementary Table 1) [11]. These criteria have been applied to many individual cases [12–14], but they have not been used to systematically identify SARS-CoV-2 reinfection in large, well-characterized cohorts.

We report clinical, laboratory, and genomic characterization of a large cohort of individuals with repeat positive NAATs ≥45 days after initial coronavirus disease 2019 (COVID-19) diagnosis at an academic medical center in Boston. We also describe the strengths and limitations of clinical, laboratory, and genomics-based assessment for the identification of SARS-CoV-2 reinfection in the clinical setting.

METHODS

Data Sources and Inclusion Criteria
We queried the Massachusetts General Hospital (MGH) electronic health record to identify subjects with at least 1 SARS-CoV-2 NAAT collected at an MGH facility ≥45 days after any prior positive NAAT (Supplementary Methods). Subjects were included in the study cohort if they had a subsequent positive NAAT collected ≥45 days after a prior positive NAAT that (1) was from a nasopharyngeal (NP) or anterior nares (AN) specimen, (2) had a cycle threshold (Ct) value <35 from any target, and (3) had residual specimen available for genomic analysis. These criteria were adapted from the CDC SARS-CoV-2 reinfection investigation criteria (Supplementary Table 1). This study was approved by the Mass General Brigham Institutional Review Board (2019P003305).

Clinical Review
We obtained demographic and clinical data regarding the initial infection episode among subjects hospitalized at MGH with COVID-19 during the first surge from 11 March 2020 to 3 June 2020 from the MGH COVID-19 Registry [15]. For these subjects' subsequent testing encounters and for all data on subjects who were not hospitalized at MGH with COVID-19 during the first surge, we extracted demographic, clinical, and laboratory data from the electronic health record into a standardized data collection form.

To assess clinical suspicion for COVID-19 reinfection, two infectious disease physician reviewers conducted detailed chart review and categorized each subject into either "low" or "moderate to high" clinical suspicion for reinfection. Reviewers were blinded to subjects' Ct values, genomic data, and other reviewers' categorizations. Features that favored "moderate to high" suspicion for reinfection included: new or recurrent upper respiratory infection symptoms, no clear alternative diagnosis, reporting a COVID-19 contact, and/or imaging and laboratory findings consistent with acute COVID-19 infection. All subjects also underwent Biofire™ extended respiratory viral testing (Supplemental Methods); however, these results were not available to reviewers unless obtained as part of clinical care. Initial clinical assessment categorizations were discordant for 6/65 (9%) subjects and were adjudicated by a third, blinded reviewer.

Ct Value Assessments
For each included subject, we reviewed all available SARS-CoV-2 NAAT results and associated Ct values. Ct values obtained from the TaqPath® COVID-19 Combo Kit were standardized to account for variability across testing platforms (Supplementary Methods and Supplementary Figure 1) [16]. Ct value assessments of reinfection were assigned to one of four categories—strongly supportive, supportive, not supportive, or inconclusive—using pre-defined criteria based on established viral kinetics of SARS-CoV-2 infection (Supplementary Table 2) [2]. We compared the yield of genomic data with specimen Ct values <30 versus ≥30 using a Fisher exact test, considering a 2-sided P value <.05 to be statistically significant.

Genomic Analysis
We performed either randomly-primed metagenomic sequencing using NexteraXT (41 samples) or tiled-amplicon sequencing with ARTICv3 primers (156 samples) on each thawed sample, sequencing the libraries as members of multiplexed pools (Supplementary Methods). Samples from most subjects did not produce complete viral genomes at all time points, which limited the ability to use only a change in PANGO lineage designation to determine reinfection. Therefore, support for reinfection was corroborated based on the following criteria, consistent with CDC guidance [11], using initial and subsequent infection period genomes: (1) well-supported phylogenetic separation (ie, change in PANGO lineage or a Shimodaira-Hasegawa likelihood ratio test node support metric >80%), (2) a higher-than-expected substitution rate for the population average (>2 substitutions/30 days), or (3) presence of a PANGO lineage during the subsequent episode that was not in local circulation (ie < 1% prevalence) during initial infection (Supplementary Methods) [11, 13, 17, 18]. Conversely, subjects whose later viral genomes descended from earlier genomes and had substitution rates less than the population average were deemed to have persistent RNA detection (ie, no reinfection) [11].

Determining the Likelihood of Reinfection

We stratified included subjects into 1 of 5 categories. Regardless of clinical or Ct value assessments, subjects with genomic demonstration of reinfection were categorized as *genomically supported reinfection*. Similarly, subjects with genomic evidence of persistent RNA detection were classified as *genomically supported persistent RNA detection* (no reinfection). If genomic analysis was indeterminate, categorization was based on clinical and Ct value assessment criteria into *probable reinfection*, *probable persistent RNA detection* (unlikely reinfection), or *inconclusive* (Supplementary Table 3).

RESULTS

Identifying Included Subjects Using Time-based Criteria

Between 14 March 2020 and 30 December 2020, 1569 individuals had at least 1 SARS-CoV-2 NAAT collected at MGH ≥45 days after a prior positive test (Figure 1A), (Supplementary Methods). Among individuals with repeat testing ≥45 days after their initial positive test, 195/1569 (12%) had a positive repeat test (Figure 1B). Of these 195 subjects, 130 (67%) were excluded because their repeat tests had a Ct value that was either ≥35 (120 subjects) or not available (6 subjects), were collected from the lower respiratory tract (3 subjects), or were negative when repeated to assess study eligibility (1 subject). In total, 65/1569 (4%) subjects with repeat testing were included in the cohort.

Yield of Genomic Analysis

SARS-CoV-2 sequencing was attempted on all available positive NP or AN specimens with a Ct <35 (n = 197) from the 65 subjects in the cohort, producing 170 partial or complete viral genome assemblies (Supplementary Figure 2; Supplementary Table 6). For 14/65 (22%) subjects, the quality of genomic data was sufficiently high to assess reinfection (Figure 1B). The 51 remaining subjects could not be genomically classified because of specimen unavailability (16/51, 31%) or insufficient genome coverage (35/51, 69%). Specimens with Ct values >30 were less likely to yield high-quality genomic data than those with Ct ≤30 (*P* value <.0001, Supplementary Figures 2 and 3), consistent with previous reports [19, 20].

Genomics-based Classification and Analysis

For the 14 subjects with sufficient genomic data at multiple time points, we identified 6 cases of genomically supported reinfection (Figure 2) and 8 cases of genomically supported persistent RNA detection (Figure 3). Of the 6 genomically supported reinfection cases, 3 (Subjects R1, R4, and R5) were classified as reinfection with moderate genomic support, as time series genome data were unavailable, but the viral PANGO lineage detected in the later timepoint had not been detected globally (R1 and R5) or was not yet in circulation in

Massachusetts (R4) during the initial time period of positivity (Figure 2), making initial infection with the lineages highly improbable. The remaining 3 cases were classified as reinfections with strong genomic support. Two cases (Subjects R3 and R6) were identified as genomically supported reinfections based on the phylogenetic separation of early and later genomes (ie, later genomes descended from other cases rather than from the earlier viral genomes); these also exhibited a greater than average nucleotide substitution rate, consistent with reinfection [11]. These 2 cases exhibited a return to the ancestral allele in later timepoints at several polymorphic positions. The last case (Subject R2) was identified based on phylogenetic separation as well as a subtle lineage change from B.1 to B.1.1. In contrast, the 8 subjects categorized as *persistent RNA detection* showed no genomic evidence for reinfection.

Among genomically supported reinfection cases, we identified mutations in the genomes from second infections relative to the ancestral reference genome (Supplementary Figure 4, Supplementary Table 4). Several of these mutations were nonsynonymous substitutions of potential functional relevance. Subject R1 had the *Mustelidae*-associated substitution S:N501T in the spike protein [21, 22], 1 also reported in a prior reinfection case [13]. Subject R6 had the substitution S:Q677H near the furin cleavage site of the spike protein [23, 24], a recurrent emergence on several distant lineages [25]. Subject R2 exhibited the substitutions N:R203K and N:G204R, observed to replicate well without cytopathy, subsequently seen in the Alpha, Gamma, and Omicron lineages, and found to increase viral particle production by elevating nucleocapsid expression [26, 27].

We characterized viral genetic changes that occurred during infection in the 8 subjects with genomically supported persistent RNA detection (Subjects P1–P8; Figure 3; Supplementary Figure 4; Supplementary Table 4) to examine viral evolution within the host. Subjects P1, P3, P4, P6, P7, and P8 exhibited viral substitutions in the middle of each time series that were not seen in the first or final timepoints, consistent with mixed viral populations and the emergence and disappearance of subclones, as previously reported [28]. The average apparent substitution rate of genomes among all subjects with genomically supported persistent RNA detection (0.9 substitutions per 30 days) was below the average rate reported for the virus across the broader population [17].

Concordance between Genomic Findings and Clinical and Ct Value Assessments

Among the 6 subjects with genomically supported reinfection, the median age was 34 years (interquartile range [IQR]: 28, 45), 5 (83%) were male, and none were immunocompromised (Table 1). Median Ct values were higher for the first NAAT compared to the subsequent qualifying NAAT (32.7 [IQR: 25.0, 33.6] vs 26.4 [IQR: 22.7, 30.8]). In 4 of 6 (67%) individuals, there was either "moderate to high" clinical suspicion or strong
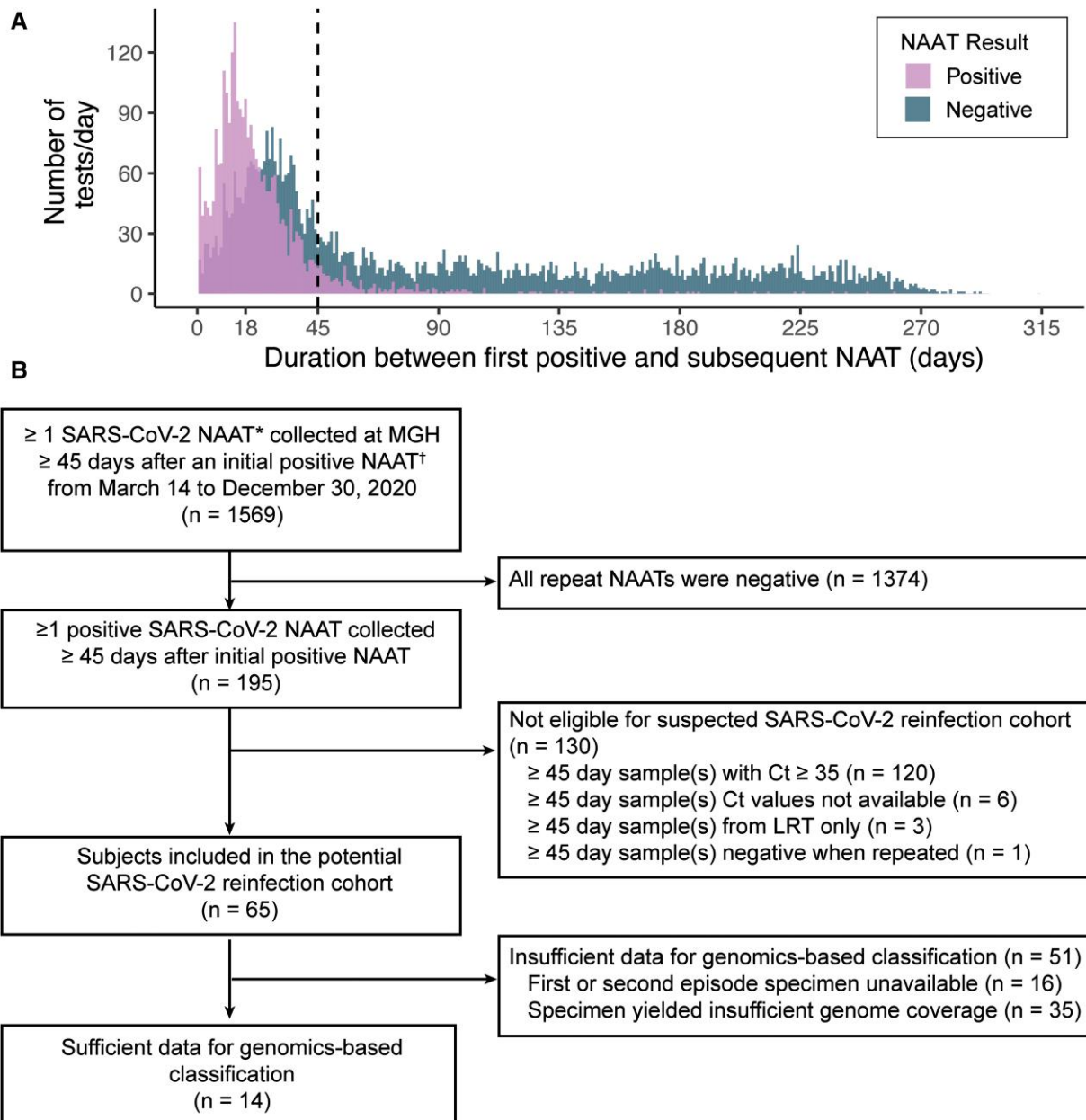
Figure 1. Identification of subjects meeting inclusion criteria. *A*, Duration between the first and subsequent positive SARS-CoV-2 NAAT for all subjects with ≥1 SARS-CoV-2 NAAT collected at the Massachusetts General Hospital ≥1 d after an initial positive SARS-CoV-2 NAAT between March 14 and 30 December 2020. *B*, Flow chart of subjects included in the study. *Positive SARS-CoV-2 NAAT includes positive/detected and presumptive positive NAAT result interpretations. †Initial positive SARS-CoV-2 NAAT includes NAATs performed at an outside facility with results available in the MGH electronic medical record. Abbreviations: Ct, cycle threshold; LRT, lower respiratory tract; MGH, Massachusetts General Hospital; NAAT, nucleic acid amplification test; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; URT, upper respiratory tract.

Ct value trajectory support for reinfection (Supplementary Figures 2 and 5), with 3 of 6 (50%) cases satisfying both criteria. Subjects R1, R4, and R6 had mild, typical COVID-19 symptoms with the second illness episode. Subject R3 had respiratory symptoms, which were attributed to post-obstructive pneumonia, and he subsequently died of lung cancer complications. Subjects R2 and R5 were asymptomatic, but Ct values were strongly suggestive of reinfection for Subject R5. For Subjects R2 and R3, clinical and Ct value assessment alone was insufficient to identify reinfection.

Among the 8 subjects with genomically supported persistent RNA detection, the median age was 40 years (IQR: 24, 64), 5 (63%) were male, and 4 (50%; Subjects P1, P2, P3, P6) were immunocompromised (Table 1). Median Ct values were lower for the first NAAT (19.2 [IQR: 16.7, 23.7]) compared to the subsequent qualifying NAAT (32.7 [IQR: 29.1, 34.2]). There was low
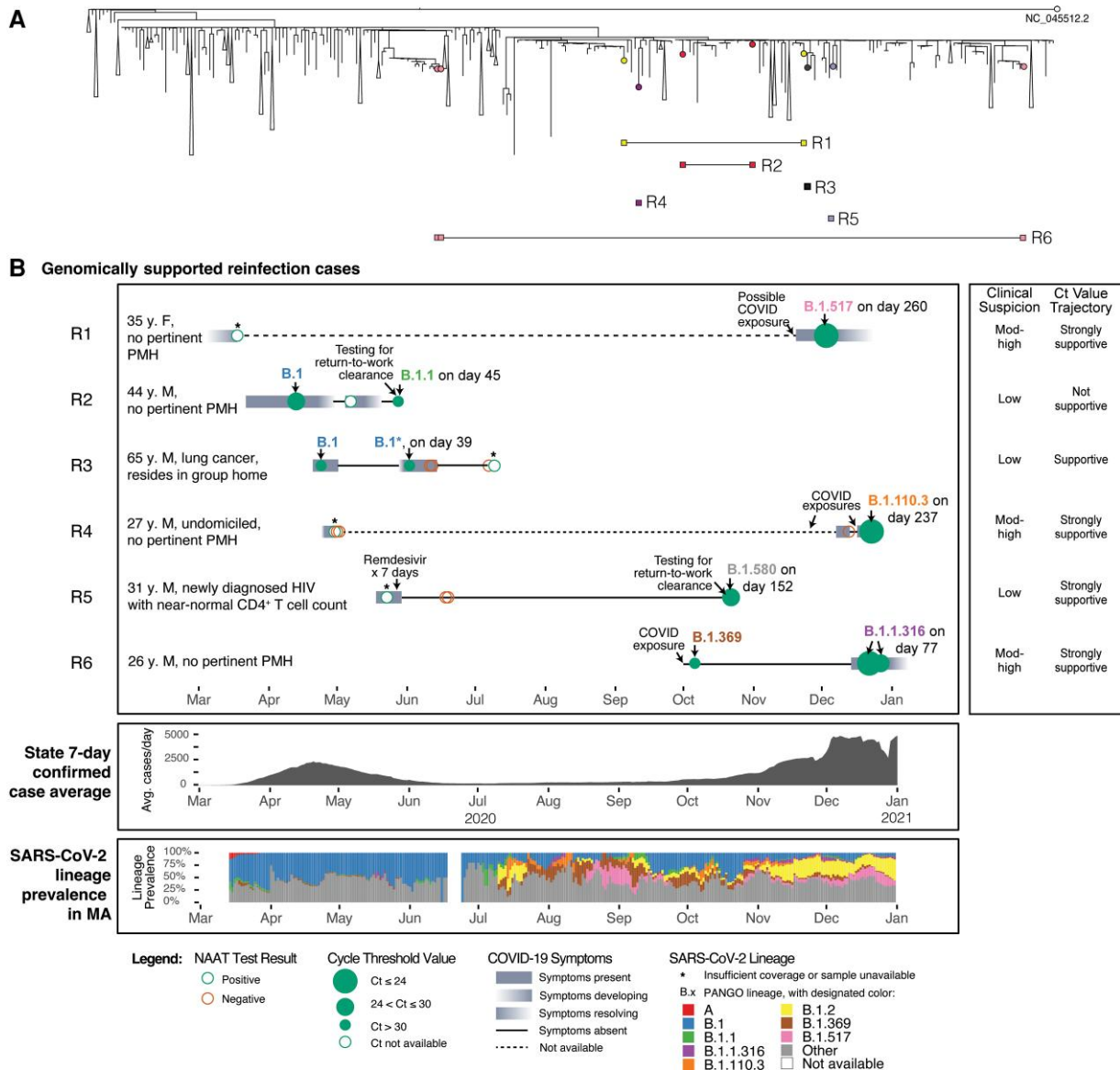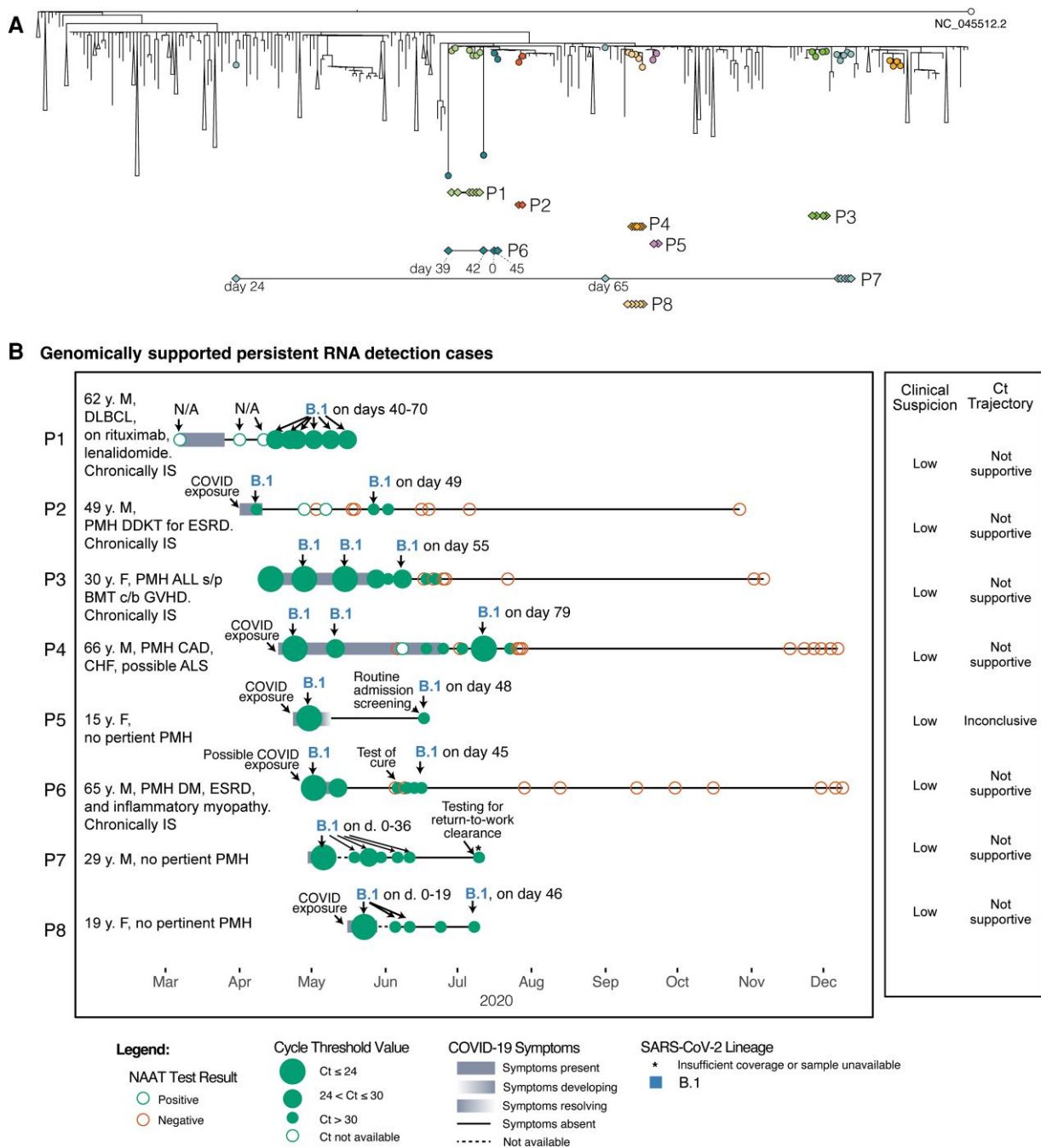
Figure 2. Comprehensive assessment of subjects with genomically supported SARS-CoV-2 reinfections in the context of MA COVID-19 cases. *A*, Global phylogeny annotated with reinfection cases. *B*, Individual timelines of subjects with genomically supported reinfection, MA COVID-19 case averages, and SARS-CoV-2 lineage prevalence in MA within the study duration. Individual timelines are annotated with nucleic acid amplification test results, Ct values (when available), COVID-19 symptomatology, and SARS-CoV-2 PANGO lineages (when available). Abbreviations: COVID, coronavirus disease; COVID-19, coronavirus disease 2019; Ct, cycle threshold; F, female; M, male; MA, Massachusetts; NAAT, nucleic acid amplification test; PMH, past medical history; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

suspicion for reinfection based on clinical and Ct value assessments in all 8 subjects (Figure 3). One immunocompromised subject receiving rituximab (Subject P1) had persistently positive NAATs with moderate Ct values (range 25–29) up to 70 days after initial infection [29].

Clinical and Ct Value Assessments Among Subjects Not Genomically Classified

When subjects could not be genomically classified (51/65 subjects, 78%), we used clinical and Ct value assessments to classify subjects as *probable reinfection, probable persistent RNA detection* (unlikely reinfection) or *inconclusive* (Supplementary Tables 3 and 5). Eight (12%) subjects were categorized as probable reinfection based on "moderate to high" clinical suspicion with a supportive Ct value assessment, or with a strongly supportive Ct value assessment alone (Supplementary Figure 6). Subjects with probable reinfection had a median duration of 87 days (IQR: 77, 111) between the first and repeat qualifying positive NAAT and lower median Ct values at repeat compared to initial testing (27.8 [IQR: 21.0, 34.1] vs 32.3 [IQR: 31.2, 32.4]; Table 1, Figure 4).

Figure 3. Comprehensive assessment of 8 subjects with genomically supported persistent RNA detection in the context of MA COVID-19 cases. *A*, Global phylogeny annotated with persistent RNA detection. *B*, Individual timelines of subjects with genomically supported persistent RNA detection annotated with nucleic acid amplification test results, Ct values (when available), COVID-19 symptomatology, and SARS-CoV-2 PANGO lineages (when available). Abbreviations: ALL, acute lymphoblastic leukemia; ALS, amyotrophic lateral sclerosis; BMT, bone marrow transplant; CAD, coronary artery disease; CHF, congestive heart failure; COVID, coronavirus disease; COVID-19, coronavirus disease 2019; Ct, cycle threshold; DDKT, deceased donor kidney transplant; DLBCL, diffuse large B-cell lymphoma; ESRD, end-stage renal disease; F, female; GVHD, graft versus host disease; IS, immunosuppressed; M, male; MA, Massachusetts; NAAT, nucleic acid amplification test; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Conversely, we classified 18 subjects (28%) as probable persistent RNA detection based on "low" clinical suspicion for reinfection and a Ct value assessment that was not supportive of reinfection (Supplementary Figure 7). These subjects had a median of 55 days (IQR: 50, 65) between the first and repeat qualifying positive NAATs, and higher median Ct values at repeat compared to initial testing (33.8 [IQR: 33.5, 34.7] vs 22.9 [IQR: 15.8, 31.2]). The remaining 25 subjects (38%) had

Table 1. Cohort Characteristics

Characteristic	Total N = 65	Genomically Supported Reinfection N = 6	Probable Reinfection[f] N = 8	Probable Persistent RNA Detection N = 18	Genomically Supported Persistent RNA Detection N = 8	Inconclusive N = 25
Male sex, n (%)	31 (48)	5 (83)	1 (13)	7 (39)	5 (63)	13 (52)
Age, y, median (IQR)	54 (31, 66)	34 (28, 45)	56 (51, 63)	55 (33, 64)	40 (24, 64)	55 (31, 69)
Race/ethnicity						
Asian, non-Hispanic	3 (5)	0 (0)	0 (0)	0 (0)	1 (13)	2 (8)
Black, non-Hispanic	11 (17)	2 (33)	1 (13)	2 (11)	2 (25)	4 (16)
Latinx/Hispanic	30 (46)	2 (33)	4 (50)	7 (39)	4 (50)	13 (52)
White, non-Hispanic	18 (28)	2 (33)	3 (38)	7 (39)	1 (13)	5 (20)
Other, non-Hispanic	2 (3)	0 (0)	0 (0)	2 (11)	0 (0)	0 (0)
Not reported	1 (2)	0 (0)	0 (0)	0 (0)	0 (0)	1 (4)
Healthcare worker with or without known exposure, n (%)	10 (15)	2 (33)	1 (13)	7 (39)	0 (0)	0 (0)
Immunocompromised, n (%)[a]	7 (11)	0 (0)	1 (13)	2 (11)	4 (50)	0 (0)
Treatment received during initial episode, n (%)						
Corticosteroids	3 (5)	0 (0)	0 (0)	1 (6)	1 (13)	1 (4)
Tocilizumab	1 (2)	0 (0)	0 (0)	1 (6)	0 (0)	0 (0)
Remdesivir	3 (5)	1 (17)	0 (0)	1 (6)	0 (0)	1 (4)
Other[b]	4 (6)	0 (0)	0 (0)	1 (6)	0 (0)	3 (12)
COVID-19 treatment received during subsequent episode, n (%)[c]	2 (3)	0 (0)	0 (0)	0 (0)	0 (0)	2 (8)
Days between first and last positive test, median (IQR)	65 (52, 89)	117 (76, 237)	87 (77, 111)	55 (50, 65)	60 (47, 70)	61 (53, 94)
Ct value of first positive test, median (IQR)	23.8 (17.5, 31.1)	32.7 (25.0, 33.6)	32.3 (31.2, 32.4)	22.9 (15.8, 31.2)	19.2 (16.7, 23.7)	22.2 (17.2, 29.1)
Ct value of subsequent positive test, median (IQR)	33.6 (31.8, 34.4)	26.4 (22.7, 30.8)	27.8 (21.0, 34.1)	33.8 (33.5, 34.7)	32.7 (29.2, 34.2)	33.8 (32.4, 34.7)
Non-COVID respiratory pathogen identified, n (%)[d]	1 (2)	0 (0)	0 (0)	1 (6)	0 (0)	0 (0)
Housing type at subsequent test, n (%)						
Private home/apartment	50 (77)	4 (67)	8 (100)	14 (78)	5 (63)	19 (76)
Shelter or experiencing homelessness	2 (3)	1 (17)	0 (0)	0 (0)	0 (0)	1 (4)
Nursing home/assisted living/rehab	8 (12)	1 (17)	0 (0)	1 (6)	1 (13)	5 (20)
Not reported	5 (8)	0 (0)	0 (0)	3 (17)	2 (25)	0 (0)
Known household contact with COVID-19, n (%)	1 (2)	0 (0)	1 (13)	0 (0)	0 (0)	0 (0)
Admitted within 28 d of subsequent qualifying episode, n (%)	20 (31)	2 (33)	1 (13)	5 (28)	2 (25)	10 (40)
Required ICU care within 28 d, n (%)	4 (6)	0 (0)	0 (0)	1 (6)	0 (0)	3 (12)
Death within 28 d, n (%)[e]	3 (5)	1 (17)	0 (0)	1 (6)	0 (0)	1 (4)

Abbreviations: COVID-19, coronavirus disease 2019; Ct, cycle threshold; ICU, intensive care unit; IQR, interquartile range.

[a]Reasons for immunocompromised state included recent bone marrow transplant or use of any of the following: oral steroids, tacrolimus, mycophenolate, rituximab, azathioprine, or lenalinamide.

[b]Patients received one or more of: hydroxychloroquine or inhaled nitric oxide.

[c]Two subjects received corticosteroids for non-COVID-19 related indications. No other COVID-19 directed therapies were administered to subjects during their subsequent episode.

[d]One patient was found to have *Bordetella pertussis* on Biofire™ (Supplementary Methods) in the setting of persistent respiratory symptoms following an initial episode of COVID-19.

[e]Cause of death included: ST-segment elevation myocardial infarction, decompensated cirrhosis, and recurrent stage IIIb non-small cell lung cancer with post-obstructive pneumonia. There were no deaths directly linked to COVID-19 infection.

[f]One individual with probable reinfection reported travel to Florida in the month prior to initial diagnosis; no other individuals had reported travel history identified by chart review.
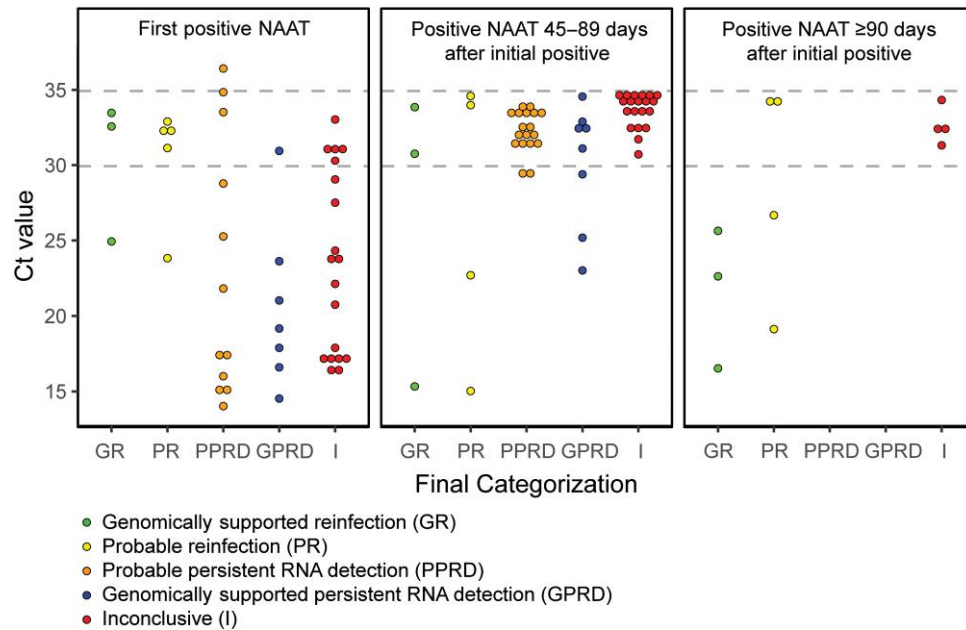
Figure 4. Comparison of Ct values between the first and subsequent (≥45 d) positive SARS-CoV-2 NAAT by final subject categorization. The lowest Ct value by final categorization for each subject is depicted. Of note, 12/65 (18%) subjects had a positive NAAT with a Ct value <35 at least 90 d after their initial positive test. Subject P4 had a day 90 NAAT with a Ct value of 32.81 but is represented in the 45–89 d group, because their lowest Ct value of 23.09 occurred on day 79. Subject I43 was excluded due to an unavailable Ct value for their day 237 positive NAAT. Abbreviations: Ct, cycle threshold; NAAT, nucleic acid amplification test; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

DISCUSSION

As SARS-CoV-2 reinfection increases in the era of Omicron and other variants of concern [7, 30], so does the importance of promptly and accurately distinguishing SARS-CoV-2 reinfection from persistent RNA detection after initial infection. Failing to recognize SARS-CoV-2 reinfection may result in missed opportunities for initiation of COVID-19 treatment, onward transmission, and delayed detection of novel SARS-CoV-2 variants [3–5, 7]. In contrast, approaching all individuals with positive NAATs long after initial infection as having reinfection places increased pressure on strained healthcare systems [31, 32] by risking misallocation of therapeutics, private hospital rooms, and personal protective equipment. Although the stakes are high, standardized approaches to distinguish reinfection from persistent RNA detection are lacking.

We systematically used clinical, laboratory, and genomic information to distinguish SARS-CoV-2 reinfection from persistent RNA detection among individuals who underwent serial SARS-CoV-2 NAATs as part of routine care. Cases of genomically supported reinfection were often mild or asymptomatic, despite generally low Ct values on repeat testing, suggesting

transmissibility [33, 34]. We found that a third (2/6) of individuals with genomically supported reinfection would have been misclassified with clinical and Ct value assessment alone, underscoring the challenge of distinguishing between SARS-CoV-2 reinfection and persistence using routinely available clinical and laboratory data. Conversely, clinical and Ct value assessment accurately identified all (8/8) individuals with genomically supported persistent RNA detection (ie, no reinfection).

We found that structured clinical and Ct value assessments performed reasonably well in distinguishing reinfection from persistent RNA detection, demonstrating 86% (12/14 subjects) concordance with viral genomic analysis, but these real-time assessments have some limitations. SARS-CoV-2 reinfection is frequently asymptomatic and may be missed by clinical assessment [35, 36]. Low or decreasing Ct values can raise suspicion for reinfection. However, Ct values vary by time from symptom onset, testing platform, and sample collection, transport, and storage conditions [37], complicating Ct value interpretation. Despite these limitations, our findings suggest that clinicians can distinguish reinfection from persistent RNA detection with reasonable accuracy by integrating clinical and Ct value assessments with established time-based definitions of reinfection.

This study demonstrates that viral genomic analysis can improve sensitivity and specificity for detecting reinfection

inconclusive results and could not be categorized due to insufficient data or discordant clinical and Ct value assessments (Supplementary Figure 8).

compared to clinical and Ct value assessment when high-quality genomic data are available from multiple SARS-CoV-2 NAATs across illness episodes. Sequencing can also provide valuable information to inform both patient- and population-level decision making [5]. Viral genomic analysis can detect amino acid substitutions that may alter treatment guidelines, (eg, suspension of bamlanivimab based on S:E484K prevalence [38, 39]). It can detect mutations evolving within immunocompromised individuals over time due to uncontrolled and potentially transmissible infection, as in subject P1 in our study [29], and enable consideration of SARS-CoV-2 retreatment. Population-level genomic surveillance may signal emergence of new variants with phenotypic differences in disease severity or transmissibility. In our study, 3 subjects with genomically supported reinfection had notable substitutions in the spike or nucleocapsid proteins [21–24, 26]. Ongoing genomic surveillance of individuals with suspected SARS-CoV-2 reinfection is important to help understand whether mutations like these promote immune escape.

Despite the potential benefits of viral genomic analysis, this study highlighted ongoing obstacles to its use for routine differentiation of SARS-CoV-2 reinfection and persistent RNA detection. First, paired specimens from multiple infection episodes suitable for genomic analysis may be unavailable. Second, recovery of full SARS-CoV-2 genomes from specimens with lower viral RNA concentrations is difficult using conventional sequencing methods [19]. We recovered sufficient genomes in less than 25% of included subjects; most subjects with successful genome recovery had Ct values <30. Third, there is currently no automated method to perform genomic analysis of potential reinfections on a timeframe fast enough to guide clinical decision making.

Overcoming these obstacles to mobilize viral genomic analysis for clinical care may be possible with substantial upfront investment. Access to sequencing and reduction in turnaround time could be achieved by embedding sequencing platforms and analyst personnel into clinical laboratories. Viral genomic yields could be improved by employing laboratory methods that reduce viral degradation (eg, minimizing dwell time and freeze-thaw cycles) or deploying shorter-amplicon sequencing [40]. Independent replicates from source material may be employed to provide additional data to enhance genome coverage and sequencing depth.

This analysis has several important limitations. First, although this is among the largest SARS-CoV-2 reinfection cohort studies to combine detailed clinical and laboratory data with viral genomic evaluation, genomic analysis distinguished potential reinfection among only a modest number of subjects. Second, our cohort represents a convenience sample of individuals undergoing repeat SARS-CoV-2 NAAT as part of routine care. As such, we cannot provide a reinfection rate; however, our results demonstrate that reinfection can occur relatively

soon after initial infection. Finally, this analysis was conducted before the emergence of SARS-CoV-2 variants of concern and deployment of COVID-19 vaccines and therapies. Although these factors may influence reinfection risk and associated characteristics, this study provides a framework for using viral genomic analysis alongside clinical and Ct value assessment that could be adapted as the COVID-19 pandemic evolves.

CONCLUSION

Although clinical and Ct value assessments were highly concordant with viral genomic analysis in identifying persistent RNA detection, they failed to identify a third of genomically supported SARS-CoV-2 reinfections. Scaling-up viral genomic analysis for real-time clinical use would likely improve detection of SARS-CoV-2 reinfections, allow for treatment among the eligible, and provide much needed insights regarding evolution of variants of concern.

References

1. Wajnberg A, Mansour M, Leven E, et al. Humoral response and PCR positivity in patients with COVID-19 in the New York City region USA: an observational study. Lancet Microbe 2020; 1:e283–9.
2. Abu-Raddad LJ, Chemaitelly H, Malek JA, et al. Assessment of the risk of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) reinfection in an intense reexposure setting. Clin Infect Dis 2021; 73:e1830–40.
3. Centers for Disease Control and Prevention. Ending isolation and precautions for people with COVID-19: interim guidance. Available at: https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html. Accessed 26 October 2021.
4. Babiker A, Marvil CE, Waggoner JJ, Collins MH, Piantadosi A. The importance and challenges of identifying SARS-CoV-2 reinfections. J Clin Microbiol 2021; 59:e02769-20.
5. Greninger AL, Dien Bard J, Colgrove RC, et al. Clinical and infection prevention applications of SARS-CoV-2 genotyping: an IDSA/ASM consensus review document. Clin Infect Dis 2021; 74:1496–502.
6. Ferrari A, Alfano G, Guaraldi G. Persistent SARS-CoV-2 positivity: an intriguing puzzle among reinfection, RNA remnants and genomic integration in COVID-19. Infect Dis Clin Pract (Baltim Md) 2021; 29:e328–9.
7. Pulliam JRC, Schalkwyk C, Govender N, et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. Science 2022; 376:eabn4947.
8. Hansen CH, Michlmayr D, Gubbels SM, Mølbak K, Ethelberg S. Assessment of protection against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: a population-level observational study. Lancet 2021; 397:1204–12.
9. Lawandi A, Warner S, Sun J, et al. Suspected SARS-CoV-2 reinfections: incidence, predictors, and healthcare use among patients at 238 U.S. healthcare facilities, June 1, 2020–February 28, 2021. Clin Infect Dis 2021; 74:1489–92.
10. Leidi A, Koegler F, Dumont R, et al. Risk of reinfection after seroconversion to SARS-CoV-2: a population-based propensity-score matched cohort study. Clin Infect Dis 2021; 74:622–9.
11. Centers for Disease Control and Prevention. Common investigation protocol for investigating suspected SARS-CoV-2 reinfection Available at: https://www.cdc.gov/coronavirus/2019-ncov/php/reinfection.html. Accessed 27 October 2020.
12. Choudhary MC, Crain CR, Qiu X, Hanage W, Li JZ. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) sequence characteristics of coronavirus disease 2019 (COVID-19) persistence and reinfection. Clin Infect Dis 2022; 74: 237–45.
13. Tomkins-Tinch CH, Daly JS, Gladden-Young A, et al. SARS-CoV-2 reinfection in a liver transplant recipient. Ann Intern Med 2021; 174:1178–80.
14. Massachi J, Donohue KC, Kelly JD. Severe acute respiratory syndrome coronavirus 2 reinfection cases corroborated by sequencing. Am J Trop Med Hyg 2021; 105:884–9.
15. Bassett IV, Triant VA, Bunda BA, et al. Massachusetts General hospital Covid-19 registry reveals two distinct populations of hospitalized patients by race and ethnicity. PLoS One 2020; 15:e0244270.
16. Rhoads D, Peaper DR, She RC, et al. College of American Pathologists (CAP) Microbiology Committee perspective: caution must be used in interpreting the cycle threshold (Ct) value. Clin Infect Dis 2021; 72:e685–6.
17. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol 2020; 6:veaa061.
18. Okoro CK, Kingsley RA, Quail MA, et al. High-resolution single nucleotide polymorphism analysis distinguishes recrudescence and reinfection in recurrent invasive nontyphoidal Salmonella typhimurium disease. Clin Infect Dis 2012; 54:955–63.
19. Lemieux JE, Siddle KJ, Shaw BM, et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science 2021; 371:eabe3261.
20. Lagerborg KA, Normandin E, Bauer MR, et al. Synthetic DNA spike-ins (SDSIs) enable sample tracking and detection of inter-sample contamination in SARS-CoV-2 sequencing workflows. Nat Microbiol 2022; 7:108–19.
21. Cai HY, Cai A. SARS-CoV2 spike protein gene variants with N501T and G142D mutation-dominated infections in mink in the United States. J Vet Diagn Invest 2021; 33:939–42.
22. Wang R, Chen J, Gao K, Wei GW. Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries. Genomics 2021; 113:2158–70.
23. Rego N, Salazar C, Paz M, et al. Emergence and spread of a B.1.1.28-derived P.6 lineage with Q675H and Q677H spike mutations in Uruguay. Viruses 2021; 13: 1801.
24. Zeng C, Evans JP, Faraone JN, et al. Neutralization of SARS-CoV-2 variants of concern harboring Q677H. mBio 2021; 12:e0251021.
25. Hodcroft EB, Domman DB, Snyder DJ, et al. Emergence in late 2020 of multiple lineages of SARS-CoV-2 Spike protein variants affecting amino acid position 677. medRxiv [Preprint]. February 21, 2021. Available from: https://doi.org/10.1101/2021.02.12.21251658
26. Syed AM, Taha TY, Tabata T, et al. Rapid assessment of SARS-CoV-2-evolved variants using virus-like particles. Science 2021; 374:1626–32.
27. Caccuri F, Zani A, Messali S, et al. A persistently replicating SARS-CoV-2 variant derived from an asymptomatic individual. J Transl Med 2020; 18:362.
28. Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature 2021; 592:277–82.
29. Choi B, Choudhary MC, Regan J, et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. N Engl J Med 2020; 383:2291–3.
30. Altarawneh HN, Chemaitelly H, Hasan MR, et al. Protection against the Omicron variant from previous SARS-CoV-2 infection. N Engl J Med 2022; 386:1288–90.
31. French G, Hulse M, Nguyen D, et al. Impact of hospital strain on excess deaths during the COVID-19 pandemic — United States, July 2020–July 2021. MMWR Morb Mortal Wkly Rep 2021; 70:1613–6.
32. U.S. Department of Health and Human Services Office of Inspector General. Hospitals reported that the COVID-19 pandemic has significantly strained health care delivery: results of a national pulse survey February 22-26, 2021: U.S. Department of Health and Human Services, 2021.
33. Jefferson T, Spencer EA, Brassey J, Heneghan C. Viral cultures for coronavirus disease 2019 infectivity assessment: a systematic review. Clin Infect Dis 2021; 73:e3884–99.
34. Singanayagam A, Patel M, Charlett A, et al. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. Euro Surveill 2020; 25:2001483.
35. Malhotra S, Mani K, Lodha R, et al. SARS-CoV-2 reinfection rate and estimated effectiveness of the inactivated whole virion vaccine BBV152 against reinfection among health care workers in New Delhi, India. JAMA Netw Open 2022; 5:e2142210.
36. Cohen C, Kleynhans J, von Gottberg A, et al. SARS-CoV-2 incidence, transmission, and reinfection in a rural and an urban setting: results of the PHIRST-C cohort study, South Africa, 2020–21. Lancet Infect Dis 2022; 22:821–34.

37. Infectious Diseases Society of America and the Association for Molecular Pathology. IDSA and AMP joint statement on the use of SARS-CoV-2 PCR cycle threshold (Ct) values for clinical decision-making. Available at: https://www.idsociety.org/globalassets/idsa/public-health/covid-19/idsa-amp-statement.pdf. Accessed 13 January 2022.

38. Jensen B, Luebke N, Feldt T, et al. Emergence of the E484K mutation in SARS-COV-2-infected immunocompromised patients treated with bamlanivi-mab in Germany. Lancet Reg Health Eur **2021**; 8:100164.

39. Peiffer-Smadja N, Bridier-Nahmias A, Ferre VM, et al. Emergence of E484K mutation following bamlanivimab monotherapy among high-risk patients infected with the alpha variant of SARS-CoV-2. Viruses **2021**; 13:1642.

40. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc **2017**; 12:1261–76.