



IMMUNOPATHOLOGY AND INFECTIOUS DISEASES

Trajectory of Growth of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variants in Houston, Texas, January through May 2021, Based on 12,476 Genome Sequences



Randall J. Olsen,^{*†} Paul A. Christensen,^{*} S. Wesley Long,^{*†} Sishir Subedi,^{*} Parsa Hodjat,^{*} Robert Olson,^{‡§} Marcus Nguyen,^{‡§} James J. Davis,^{‡§} Prasanti Yerramilli,^{*} Matthew O. Saavedra,^{*} Layne Pruitt,^{*} Kristina Repond,^{*} Madison N. Shyer,^{*} Jessica Cambric,^{*} Ryan Gadd,^{*} Rashi M. Thakur,^{*} Akanksha Batajoo,^{*} Ilya J. Finkelstein,[¶] Jimmy Gollihar,^{*||} and James M. Musser^{*†}

From the Center for Molecular and Translational Human Infectious Diseases Research,^{*} Department of Pathology and Genomic Medicine, Houston Methodist Research Institute and Houston Methodist Hospital, Houston, Texas; the Departments of Pathology and Laboratory Medicine, and Microbiology and Immunology,[†] Weill Cornell Medical College, New York, New York; the Consortium for Advanced Science and Engineering,[‡] University of Chicago, Chicago, Illinois; Computing, Environment and Life Sciences,[§] Argonne National Laboratory, Lemont, Illinois; the Department of Molecular Biosciences and Institute of Molecular Biosciences,[¶] The University of Texas at Austin, Austin, Texas; and the Combat Capabilities Development Command (CCDC) Army Research Laboratory-South,^{||} University of Texas, Austin, Texas

Accepted for publication
July 2, 2021.

Address correspondence to
James M. Musser, M.D., Ph.D.,
Department of Pathology and
Genomic Medicine, Houston
Methodist Research Institute,
6565 Fannin St., Ste. B490,
Houston, TX 77030. E-mail:
jmmusser@houstonmethodist.org.

Certain genetic variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are of substantial concern because they may be more transmissible or detrimentally alter the pandemic course and disease features in individual patients. SARS-CoV-2 genome sequences from 12,476 patients in the Houston Methodist health care system diagnosed from January 1 through May 31, 2021 are reported here. Prevalence of the B.1.1.7 (Alpha) variant increased rapidly and caused 63% to 90% of new cases in the latter half of May. Eleven B.1.1.7 genomes had an E484K replacement in spike protein, a change also identified in other SARS-CoV-2 lineages. Compared with non-B.1.1.7-infected patients, individuals with B.1.1.7 had a significantly lower cycle threshold (a proxy for higher virus load) and significantly higher hospitalization rate. Other variants [eg, B.1.429 and B.1.427 (Epsilon), P.1 (Gamma), P.2 (Zeta), and R.1] also increased rapidly, although the magnitude was less than that in B.1.1.7. Twenty-two patients infected with B.1.617.1 (Kappa) or B.1.617.2 (Delta) variants had a high rate of hospitalization. Breakthrough cases ($n = 207$) in fully vaccinated patients were caused by a heterogeneous array of virus genotypes, including many not currently designated variants of interest or concern. In the aggregate, this study delineates the trajectory of SARS-CoV-2 variants circulating in a major metropolitan area, documents B.1.1.7 as the major cause of new cases in Houston, TX, and heralds the arrival of B.1.617 variants in the metroplex. (*Am J Pathol* 2021; 191: 1754–1773; <https://doi.org/10.1016/j.ajpath.2021.07.002>)

The global pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that began in early 2020 has been challenging for every academic health center and health system, hospital, and public health system in the United States and countries worldwide.^{1–7} The pandemic has also provided unprecedented opportunities for basic and translational research in all biomedical fields. Molecular population genomics of SARS-CoV-2 were systematically

analyzed in the ethnically and socioeconomically diverse metropolitan Houston, TX, area (population 7 million) since

Supported by the Houston Methodist Academic Institute Infectious Diseases Fund; and the National Institute of Allergy and Infectious Diseases, NIH, Department of Health and Human Services contract 75N93019C00076 (J.J.D. and R.O.).

R.J.O., P.A.C., S.W.L., and S.S. contributed equally to this work.
Disclosures: None declared.

the first coronavirus disease 2019 (COVID-19) cases were reported in early March 2020.^{8–11} These studies are facilitated by a central molecular diagnostic laboratory that comprehensively identifies and retains all COVID-19 diagnostic specimens from our large health care system, which includes eight hospitals, emergency care clinics, and outpatient centers distributed throughout the metropolitan region. In addition, the longstanding interest in pathogen genomics and sequencing infrastructure was leveraged to investigate the spread of SARS-CoV-2 in metropolitan Houston.^{8–16} SARS-CoV-2 viruses causing infections in the earliest phase of the pandemic affecting Houston had substantial genomic diversity and are progeny of strains derived from several continents, including Europe and Asia.^{8,9} These findings indicated that SARS-CoV-2 was introduced into our region many times independently by individuals who had traveled from different parts of the country and the world. Subsequently, sequence analysis of 5085 genomes causing the first disease wave and massive second disease wave in Houston showed that all strains in the second wave had a D614G amino acid replacement in the spike protein.⁹ The D614G polymorphism increases human transmission and infectivity *in vitro* and *in vivo* in animal infection models.^{17–22} More importantly, this was the first study to analyze the molecular architecture of SARS-CoV-2 in two infection waves in any major metropolitan region.

One of the key goals since the start of the pandemic has been to sequence all positive SARS-CoV-2 specimens from patients in our hospital system and rapidly identify mutations that may be associated with detrimental patient outcome, including therapeutic or vaccine failure. Similarly, with the recognition of an increasing number of SARS-CoV-2 variants of interest (VOIs) and variants of concern (VOCs) by public health agencies, such as the US CDC, World Health Organization, and Public Health England (<https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html>, last accessed June 8, 2021; and <https://www.gov.uk/government/collections/new-sars-cov-2-variant>, last accessed June 8, 2021), there is now substantial domestic and international need to identify these virus genotypes rapidly and understand their velocity and patterns of dissemination. In particular, VOC B.1.1.7 (also termed Alpha), first identified in the United Kingdom, is of special interest because it has the ability to transmit effectively, it can spread through populations rapidly, and has been reported to have a significantly higher mortality rate than non-B.1.1.7 infections (Virological, <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spoke-mutations/563>, last accessed June 8, 2021; Public Health England, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/947048/technical_briefing_voc_sh_njl2_sh2.pdf, last accessed June 8, 2021; New and Emerging Respiratory Virus Threats Advisory Group, <https://app.box.com/s/3lkcbxepqixkg4mv640dpv>

vg978ixjtf/file/756963730457, last accessed June 8, 2021; Centre for Mathematical Modelling of Infectious Diseases, <https://cmmid.github.io/topics/covid19/uk-novel-variant.html>, last accessed June 8, 2021; and <https://virological.org/t/lineage-specific-growth-of-sars-cov-2-b-1-1-7-during-the-english-national-lockdown/575>, last accessed June 8, 2021).^{23–36} VOCs B.1.351 (β) and P.1 (Gamma), found to cause widespread disease in South Africa and Brazil, respectively, have sequence changes in spike protein that make them less susceptible to host and some therapeutic antibodies.^{37–40} Recently, two additional VOIs, B.1.427 and B.1.429 (Epsilon), were recognized by the CDC in part because of their rapid transmission in many California communities⁴¹ (Outbreak.info, <https://outbreak.info/situation-reports?pango=b.1.427>, last accessed June 8, 2021; and <https://outbreak.info/situation-reports?pango=b.1.4279>, last accessed June 8, 2021).

Based on sequencing 20,453 SARS-CoV-2 genomes causing COVID-19 disease in Houston, all named VOIs and VOCs are circulating in the metropolitan region, making it the first community to document their presence.¹⁰ A follow-up study reported rapid increase of VOC UK B.1.1.7 in Houston¹¹; cases infected with the variant were estimated to have a doubling time of approximately 7 days. This rapid B.1.1.7 growth trajectory raised the possibility that this variant would cause nearly all new COVID-19 cases in metropolitan Houston by the end of March or early April 2021, a time frame similar to an estimate made in late January by the CDC.³³

This study reports integrated virus genome and patient data for 12,476 unique COVID-19 cases identified between January 1, 2021, and May 31, 2021, including 3276 patients with the B.1.1.7 VOC. In the latter half of May, depending on the day, 63% to 90% of all new COVID-19 cases in metropolitan Houston were caused by B.1.1.7. Linked medical record information, available for virtually all sequenced genomes, was used to study the relationship between virus genotypes and patient phenotypes. Patients infected with B.1.1.7 had significantly lower cycle threshold (C_T) values in nasopharyngeal specimens (considered to be a proxy for higher virus load) and a significantly higher hospitalization rate compared with non-B.1.1.7 patients. There was no difference between these two groups in hospital length of stay or mortality. Of the 3276 B.1.1.7 genomes, 11 (0.3%) had an E484K change in spike protein that reduces binding by some neutralizing antibodies. Unexpectedly, five cases of B.1.1.7 were detected from samples collected in early December, resulting in a revised time frame for the introduction of this variant to Houston. Twenty-two patients were identified with COVID-19 caused by B.1.617.1 (Kappa) or B.1.617.2 (Delta) variants reported to be causing widespread disease and extensive public health problems in India, other Southeast Asian countries, and many regions of the United Kingdom (World Health Organization, <https://www.who.int/publications/m/item/>

[weekly-epidemiological-update-on-covid-19—8-june-2021](#), last accessed June 9, 2021).^{42–49} These patients also had a high rate of hospitalization. Vaccine breakthrough cases ($n = 207$) were caused by diverse virus genotypes, many of which were not VOCs or VOIs. Our genome data show that VOCs and VOIs now account for the great majority of all new COVID-19 cases in this region, identify B.1.1.7 as the major cause of new cases in Houston, and document the arrival and spread of B.1.617 variants in the Houston metroplex.

Materials and Methods

Patient Specimens

Specimens were obtained from registered patients at Houston Methodist hospitals, associated facilities (eg, urgent care centers), and institutions in the Houston metropolitan region that use our laboratory services. Virtually all individuals had signs or symptoms consistent with COVID-19 disease. We analyzed a comprehensive sample of genomes obtained from January 1, 2021, through May 31, 2021. This time frame was chosen for convenience because it represents the period during which, at the onset of the study, we identified an uptick in identification of VOIs and VOCs. The study included all 12,476 unique patients identified in this time frame. The work was approved by the Houston Methodist Research Institute Institutional Review Board (IRB1010-0199).

SARS-CoV-2 Molecular Diagnostic Testing

Specimens obtained from symptomatic patients with a suspicion for COVID-19 disease were tested in the Molecular Diagnostics Laboratory at Houston Methodist Hospital using assays granted Emergency Use Authorization from the US Food and Drug Administration ([https://www.fda.gov/medical-devices/emergency-situations-medical-devices/faqs-diagnostic-testing-sars-cov-2#offeringtests](#), last accessed June 7, 2021). As a hedge against supply chain strictures, multiple molecular testing platforms were used, including the COVID-19 test or RP2.1 test with BioFire Film Array instruments (BioFire Diagnostics, Salt Lake City, Utah), the Xpert Xpress SARS-CoV-2 test using Cepheid GeneXpert Infinity or Cepheid GeneXpert Xpress IV instruments (Cepheid, Sunnyvale, CA), the cobas SARS-CoV-2 & Influenza A/B Assay using the Roche Liat system (Roche Diagnostics, Indianapolis, IN), the SARS-CoV-2 Assay using the Hologic Panther instrument (Hologic, Marlborough, MA), the Aptima SARS-CoV-2 Assay using the Hologic Panther Fusion system (Hologic, Marlborough, MA), the Cobas SARS-CoV-2 test using the Roche 6800 system (Roche Diagnostics, Indianapolis, IN), and the SARS-CoV-2 assay using Abbott Alinity m instruments (Abbott, Abbott Park, IL). The great majority of tests were performed

on material obtained from nasopharyngeal swabs immersed in universal transport media; oropharyngeal or nasal swabs, bronchoalveolar lavage fluid, or sputum treated with dithiothreitol was sometimes used. Standardized specimen collection methods were used (Vimeo, [https://vimeo.com/396996468/2228335d56](#), last accessed June 7, 2021).

SARS-CoV-2 Genome Sequencing

Libraries for whole virus genome sequencing were prepared according to version 3 of the ARTIC nCoV-2019 sequencing protocol (Artic Network, [https://artic.network/ncov-2019](#), last accessed June 7, 2021). We used a semi-automated workflow that employed BioMek i7 liquid handling workstations (Beckman Coulter Life Sciences, Indianapolis, IN) and MANTIS automated liquid handlers (FORMULATRIX, Bedford, MA). Short sequence reads were generated with a NovaSeq 6000 instrument (Illumina, San Diego, CA). For continuity of the epidemiologic analysis in the study period, some genome sequences reported in a recent publication were included.¹⁰

SARS-CoV-2 Genome Sequence Analysis and Identification of Variants

Viral genomes were assembled with the BV-BRC SARS-CoV2 assembly service (The Bacterial and Viral Bioinformatics Resource Center, [https://www.bv-brc.org/app/comprehensivesars2analysis](#), last accessed June 7, 2021, requires registration). The One Codex SARS-CoV-2 variant calling and consensus assembly pipeline was used to assemble all sequences (GitHub, [https://github.com/onecodex/sars-cov-2.git](#), last accessed June 7, 2021) using default parameters and a minimum read depth of three. Briefly, the pipeline uses seqtk version 1.3-r116 for sequence trimming (GitHub, [https://github.com/lh3/seqtk.git](#), last accessed June 7, 2021); minimap version 2.1 ([https://github.com/lh3/minimap2](#), last accessed August 09, 2021) for aligning reads against reference genome Wuhan-Hu-1 ([https://www.ncbi.nlm.nih.gov/nuccore/1798174254](#); NC_045512.2); samtools version 1.11 ([http://www.htslib.org](#), last accessed August 09, 2021) for sequence and file manipulation; and iVar version 1.2.2 ([https://github.com/andersen-lab iVar/releases](#), last accessed August 09, 2021) for primer trimming and variant calling. Genetic lineages, VOCs, and VOIs were identified on the basis of genome sequence data and designated by Pangolin version 3.0.5 ([https://github.com/cov-lineages/pangoLEARN](#), last accessed August 09, 2021) with pangoLEARN module 2021 to 06-05 (SARS-CoV-2 lineages, [https://cov-lineages.org/pangolin.html](#), last accessed June 7, 2021).

Table 1 Summary of Pertinent Patient Metadata for the 12,476 Unique Patients

Variable	B.1.1.7 variant	Other variants	Total	Statistical analysis
Patients with data	3276 (26.3)	9200 (73.7)	12,476	
Patient characteristics				
Median age, years	49.9	53.7	52.5	$P < 0.0001$ (<i>U</i> -test)
Female sex	1693 (51.7)	4880 (53.0)	6573 (52.7)	$P = 0.1854$
Male sex	1583 (48.3)	4320 (47.0)	5903 (47.3)	(Fisher exact test)
Race/ethnicity				
White	1304 (40.1)	3720 (40.6)	5024 (40.5)	$P < 0.0001$
Hispanic or Latino	945 (29.1)	2720 (29.7)	3665 (29.5)	(χ^2 test)
Black	748 (23.0)	1644 (18.0)	2392 (19.3)	
Asian	128 (3.9)	549 (6.0)	677 (5.5)	
Native American	16 (0.5)	28 (0.3)	44 (0.4)	
Hawaiian/Pacific Islander	3 (0.1)	23 (0.3)	26 (0.2)	
Unavailable	107 (3.3)	472 (5.2)	579 (4.7)	
BMI, kg/m ²				
Median	30.4	29.5	<i>n</i> = 11,009	$P < 0.0001$ (<i>U</i> -test)
<18.5	48 (1.5)	140 (1.5)	188 (1.5)	$P < 0.0001$
18.5–25	506 (15.4)	1608 (17.5)	2114 (16.9)	(χ^2 test)
25–30	857 (26.2)	2520 (27.4)	3377 (27.1)	
30–35	740 (22.6)	1833 (19.9)	2573 (20.6)	
≥35	811 (24.8)	1946 (21.2)	2757 (22.1)	
Unknown	314 (9.6)	1153 (12.5)	1467 (11.8)	
Admission data				
Admitted	1768 (54.0)	4265 (46.4)	6033 (48.4)	$P < 0.0001$
Not admitted	1508 (46.0)	4935 (53.6)	6443 (51.6)	(Fisher exact test)
Median LOS, days	5.1	5.2	5.2	Odds ratio: 1.357 (95% CI, 1.252–1.469) $P = 0.8917$ (<i>U</i> -test)
Maximum respiratory support				
ECMO	8 (0.5)	17 (0.4)	25 (0.4)	$P = 0.0135$
Mechanical ventilation	150 (8.5)	365 (8.6)	515 (8.5)	(χ^2 test)
Noninvasive ventilation	169 (9.6)	433 (10.2)	602 (10.0)	
High-flow oxygen	358 (20.2)	696 (16.3)	1054 (17.5)	
Low-flow oxygen	734 (41.5)	1830 (42.9)	2564 (42.5)	
Room air	349 (19.7)	924 (21.7)	1273 (21.1)	
Mortality				
Alive	3132 (95.6)	8760 (95.2)	11,892 (95.3)	$P = 0.3862$
Deceased	144 (4.4)	440 (4.8)	584 (4.7)	(Fisher exact test)
				Odds ratio: 0.915 (95% CI, 0.755–1.111)
Median PCR C _T				
Abbott Alinity	23.9 (<i>n</i> = 1133)	26.8 (<i>n</i> = 3344)	<i>n</i> = 4477	$P < 0.0001$ (<i>U</i> -test)
Hologic Panther	25.0 (<i>n</i> = 385)	26.2 (<i>n</i> = 1574)	<i>n</i> = 1959	$P = 0.0274$ (<i>U</i> -test)
Vaccine				
No vaccine	3023 (92.3)	8715 (94.7)	11,738 (94.1)	$P < 0.0001$
>7 Days past first vaccine	127 (3.9)	404 (4.4)	531 (4.3)	(χ^2 test)
>14 Days past second vaccine	126 (3.8)	81 (0.9)	207 (1.7)	

Data are given as *n* (%), unless otherwise indicated. Data include 12,476 unique patients with high-quality sequence results between January 1, 2021, and May 31, 2021.

BMI, body mass index; ECMO, extracorporeal membrane oxygenation; LOS, length of stay.

Patient Metadata and Geospatial Analysis

Patient metadata (Tables 1 and 2) were acquired from the electronic medical record by standard informatics methods.

Patient home address zip codes were used to visualize the geospatial distribution of spread for each VOC and VOI. Figures were generated with Tableau version 2020.3.4 (Tableau Software, LLC, Seattle, WA). A vaccination

Table 2 Summary of Pertinent Patient Metadata for B.1.617.1/B.1.617.2 Patients (Excluding B.1.1.7 Patients)

Variable	B.1.617 variants	Other variants	Total	Statistical analysis
Patients with data	22 (0.2)	9178 (99.8)	9200	
Race/ethnicity				
Asian	7 (33.3)	542 (5.9)	549 (6.0)	$P < 0.001$
White	8 (38.1)	3712 (40.6)	3720 (40.6)	(Fisher exact test) (Asian versus non-Asian)
Hispanic or Latino	3 (14.3)	2717 (29.7)	2720 (29.7)	Odds ratio:
Black	0 (0)	1644 (18.0)	1644 (18.0)	7.927 (95% CI, 2.973–19.40)
Native American	1 (4.8)	27 (0.3)	28 (0.3)	
Hawaiian/Pacific Islander	1 (4.8)	22 (0.2)	23 (0.3)	
Unavailable	1 (4.8)	471 (5.2)	472 (5.2)	
Admission data				
Admitted	17 (77.3)	4248 (46.3)	4265 (46.4)	$P = 0.0045$
Not admitted	5 (22.7)	4930 (53.7)	4935 (53.6)	(Fisher exact test)
Median LOS, days	6.7	5.2		Odds ratio: 3.946 (95% CI, 1.529–9.788)
				$P = 0.5388$
				(<i>U</i> -test)
Highest level of care				
ICU	3 (17.6)	639 (15.0)	642 (15.1)	$P = 0.5474$
IMU	1 (5.9)	93 (2.2)	94 (2.2)	(<i>U</i> -test)
Other inpatient	13 (76.5)	3516 (82.8)	3529 (82.7)	
Maximum respiratory support				
ECMO	0 (0)	17 (0.4)	17 (0.4)	$P > 0.9999$
Mechanical ventilation	3 (17.6)	362 (8.5)	365 (8.6)	(Fisher exact test) (room air and low-flow
Noninvasive ventilation	1 (5.9)	432 (10.2)	433 (10.2)	oxygen versus other)
High-flow oxygen	2 (11.8)	694 (16.3)	696 (16.3)	Odds ratio: 1.006 (95% CI, 0.3803–2.647)
Low-flow oxygen	8 (47.1)	1822 (42.9)	1830 (42.9)	
Room air	3 (17.6)	921 (21.7)	924 (21.7)	
Mortality				
Alive	20 (90.9)	8740 (95.2)	8760 (95.2)	$P = 0.2838$
Deceased	2 (9.1)	438 (4.8)	440 (4.8)	(Fisher exact test)
				Odds ratio: 1.995 (95% CI, 0.459–7.462)

Data are given as *n* (%), unless otherwise indicated. Data include 9200 unique patients with high-quality sequence results between January 1, 2021, and May 31, 2021.

ECMO, extracorporeal membrane oxygenation; ICU, intensive care unit; IMU, intermediate care unit; LOS, length of stay.

breakthrough case was defined as a PCR-positive sample from a symptomatic patient obtained >14 days after full vaccination (ie, both doses of the Pfizer or Moderna mRNA vaccines) was completed.

Results

Epidemiologic Trajectory and Patient Overview

Metropolitan Houston has experienced three distinct epidemiologic peaks of COVID-19 (Figure 1). The timing and shape of the epidemiologic curve for Houston Methodist patients mirrors the curve for the metropolitan region (Harris County Public Health, <https://covid-harriscounty.hub.arcgis.com/pages/cumulative-data>, last accessed June 7, 2021). The third wave of COVID-19 started in approximately early November, following a prolonged disease trough occurring after the second wave (Figure 1). Approximately 12,476 patients were studied from January 1, 2021, through May 31, 2021, a period during which most of the VOIs and VOCs were

initially identified in Houston, and several of them increased substantially (Table 1 and Figures 2, 3, and 4).

The median age of the patients studied was 52.5 years, and 53% were female; 6033 (48.4%) patients required hospitalization. The ethnic distribution of the patients (Table 1) broadly reflects metropolitan Houston, which has a majority-minority population composition. Median length of stay was 5.2 days, and the 28-day mortality rate was 4.7%.

Occurrence of VOIs and VOCs

The CDC has identified nine VOIs [B.1.427 and B.1.429 (Epsilon), B.1.525 (also termed Eta), B.1.526 and B.1.526.1 (Iota), P.2 (Zeta), B.1.617, B.1.617.1 (Kappa), and B.1.617.3] and four VOCs [B.1.1.7 (Alpha), B.1.617.2 (Delta), P.1 (Gamma), and B.1.351 (Beta)] based on heightened concern about potential or proven threat to public health and individual patients. The following VOIs were identified in our comprehensive sample of 12,476 genome sequences: B.1.427 (*n* = 58), B.1.429 (*n* = 293), B.1.525 (*n* = 26), B.1.526 (*n* = 39), B.1.526.1 (*n* = 9),

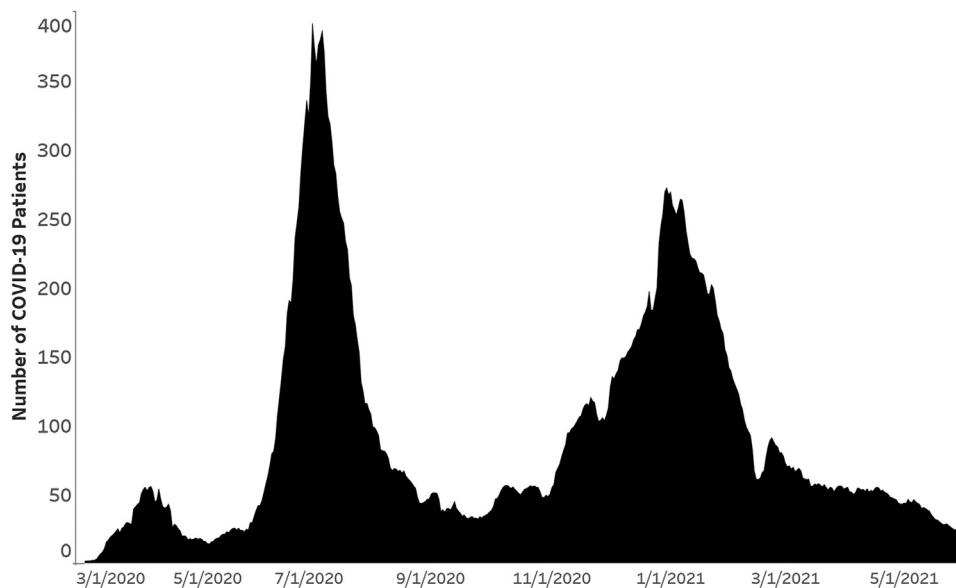


Figure 1 Epidemiologic curve showing three waves of SARS-CoV-2 infections in Houston Methodist patients. Daily totals are shown as a \pm 3-day moving average. The figure was generated with Tableau version 2020.3.4.

B.1.617.1 ($n = 5$), and P.2 ($n = 74$). All four VOCs were found, including B.1.1.7 ($n = 3276$), P.1 ($n = 87$), B.1.617.2 ($n = 17$), and B.1.351 ($n = 4$) (Figures 2, 3, 4, and 5). B.1.1.7 rapidly increased in the population and now dominates the new-infection landscape in Houston (Figures 2, 3, and 4). In the last half of May, depending on the specific day, the B.1.1.7 variant caused 63% to 90% of new COVID-19 cases. In addition, cases caused by variants P.1, P.2, B.1.429, and the B.1.617 family also increased during the study period, although not to the magnitude of B.1.1.7 infections (Figures 2, 3, and 4).

Variants Genetically Related to B.1.617

Although comprehensive data are not available from India, the B.1.617, B.1.617.1, B.1.617.2, and B.1.617.3 variants were recently described as causing widespread COVID-19 disease in that country^{42–44} (World Health Organization, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—11-may-2021>, last accessed May 16, 2021) and have been designated as VOIs or VOCs by the CDC. B.1.617-family variants also have been reported to be prominent causes of new COVID-19 cases in other countries in Southeast Asia and the United Kingdom (World Health Organization, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—8-june-2021>, last accessed June 9, 2021).^{42–49} Variant B.1.617 is resistant to the monoclonal antibody bamlanivimab (LY-Cov555), as assessed by an *in vitro* host-cell entry assay,⁴⁵ and B.1.617.1 has been reported to be highly virulent in hamsters following intranasal inoculation.⁴⁴ These two variants are characterized by a core group of amino acid replacements in spike protein: L452R, T478K or E484Q, D614G, and P681R (Figure 6). More importantly, genetic

variation exists among sequences classified as B.1.617.1 and B.1.617.2. Among these 5 B.1.617.1 and 17 B.1.617.2 variant samples, 4 and 11 distinct subvariants were identified, respectively (Figure 6). Two of the patients with B.1.617.1 and three of the patients with B.1.617.2 had a recent travel history to a high-prevalence country. One additional B.1.617.2 patient had a history of recent international travel to an unspecified country. Examination of the metadata available for patients with B.1.617 variants found that relative to non-B.1.617 patients, a higher percentage of cases were of Asian ethnicity, and a lower percentage of patients were Hispanic or Latino (Table 2). In addition, B.1.617 patients had a higher hospitalization rate than non-B.1.617 patients (Table 2).

C_T Value Comparison of B.1.1.7 and Non-B.1.1.7 Samples

Early in the pandemic, it was reported that nasopharyngeal samples from patients infected with strains having the spike protein D614G variant have, on average, significantly lower C_T values (considered to be a proxy for higher virus loads) on initial diagnosis.^{9,17} Most authorities think that higher virus load in the upper respiratory tract is related to enhanced ability to spread and infect others, although there are many factors that contribute to virus transmission and disease. The hypothesis that specimens from patients with B.1.1.7 infections had lower C_T values compared with non-B.1.1.7 patients based on data generated by the Abbott Alinity m or Hologic Panther molecular diagnostic assays was tested. Consistent with the hypothesis, patient samples with the B.1.1.7 variant had significantly lower mean C_T value (Table 1 and Figure 7) on these instruments, and thus likely have higher nasopharyngeal virus loads. Next, the

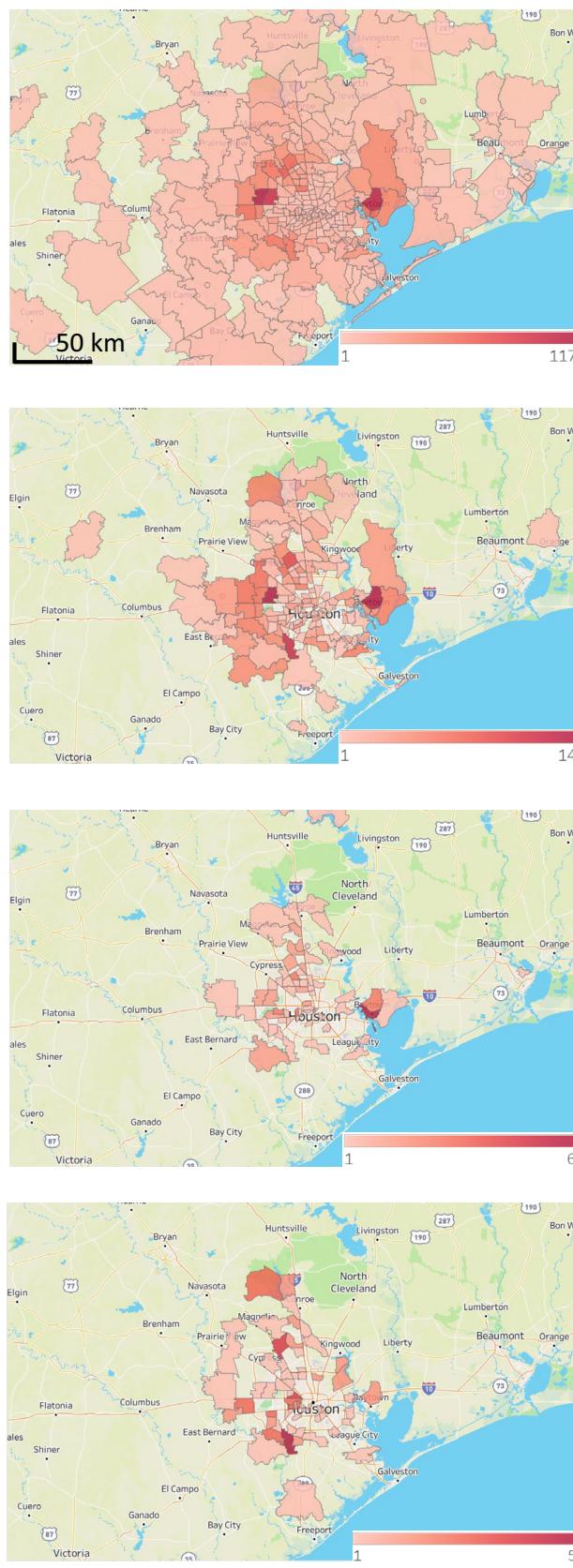
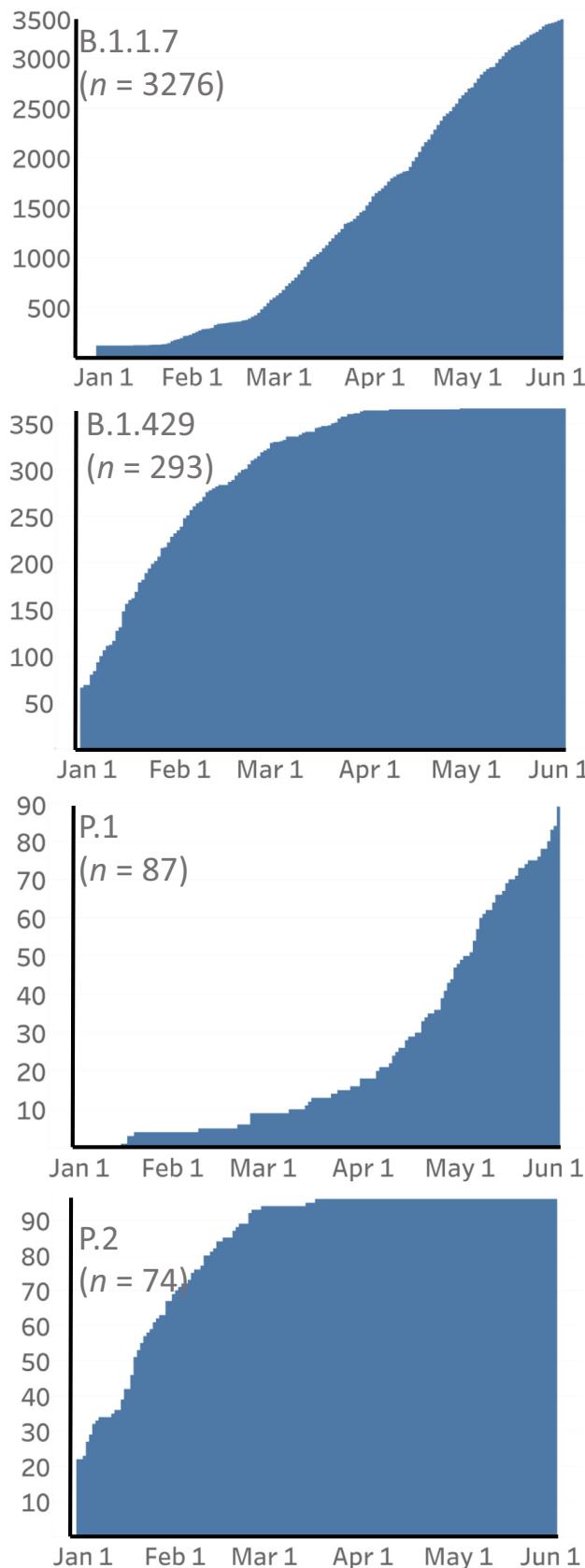


Figure 2 Cumulative increase in SARS-CoV-2 B.1.1.7, B.1.429, P.1, and P.2 variants and their distribution in metropolitan Houston, TX. The time frame used is January 1, 2021, through May 31, 2021. **Left column:** The cumulative increase in unique patients with each variant. **Right column:** The geospatial distribution of these variants based on the home address zip code for each patient. Figures were generated with Tableau version 2020.3.4.

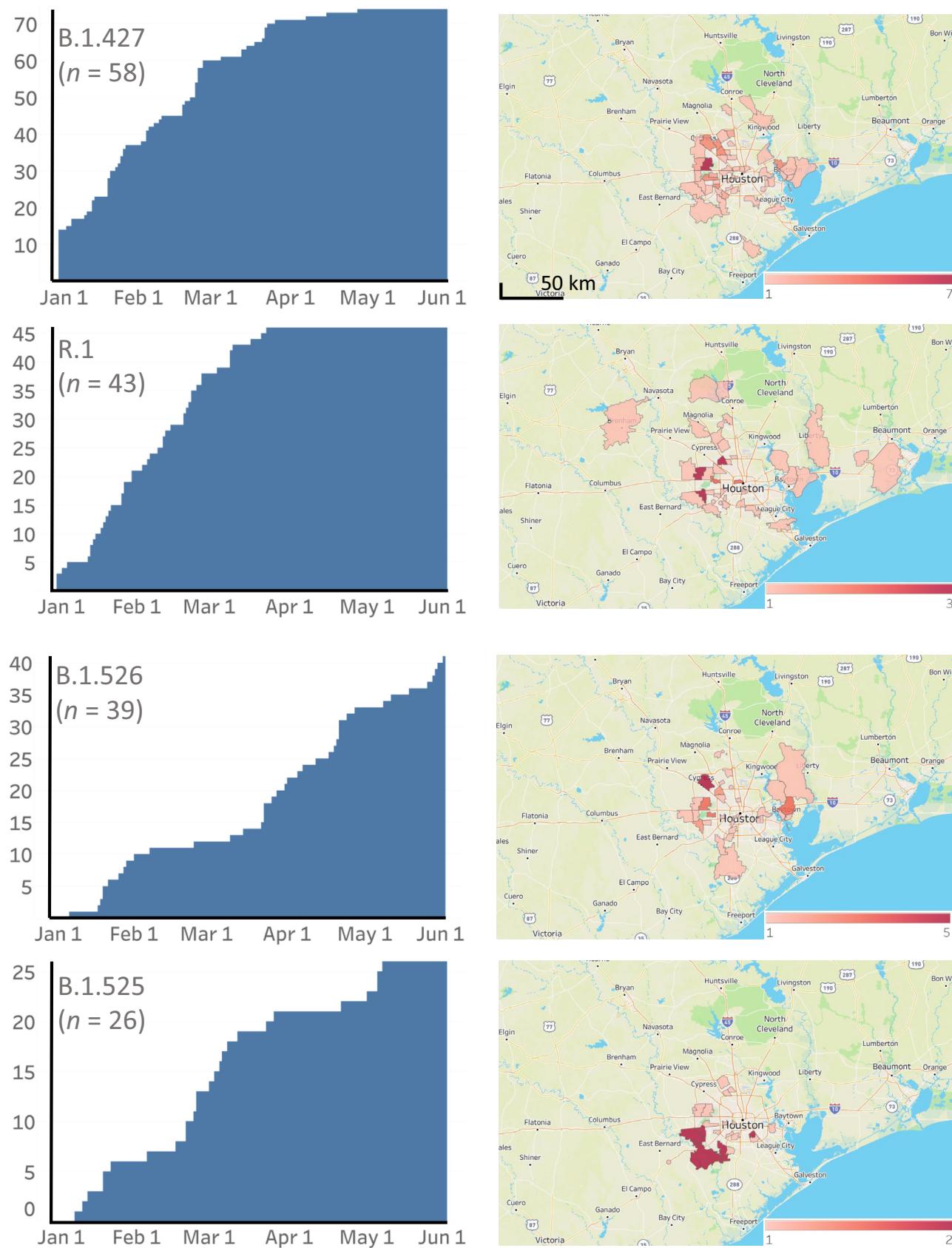


Figure 3 Cumulative increase in SARS-CoV-2 B.1.427, R.1, B.1.526, and B.1.525 variants and their distribution in metropolitan Houston, TX. The time frame used is January 1, 2021, through May 31, 2021. **Left column:** The cumulative increase in unique patients with each variant. **Right column:** The geospatial distribution of these variants based on the home address zip code for each patient. Figures were generated with Tableau version 2020.3.4.

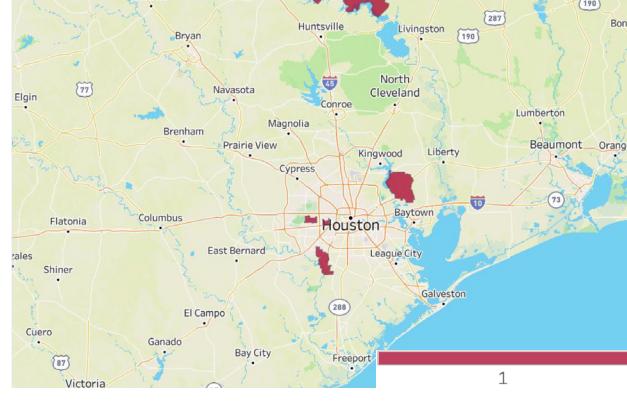
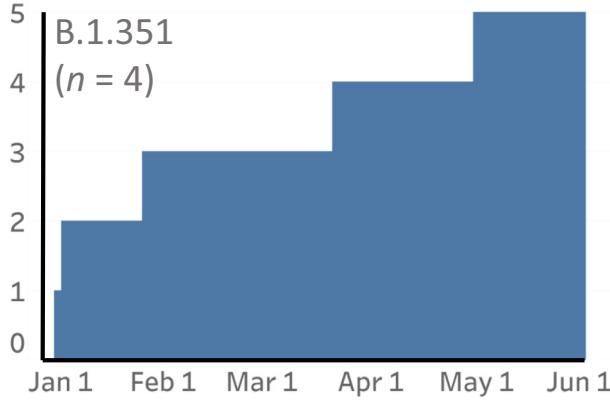
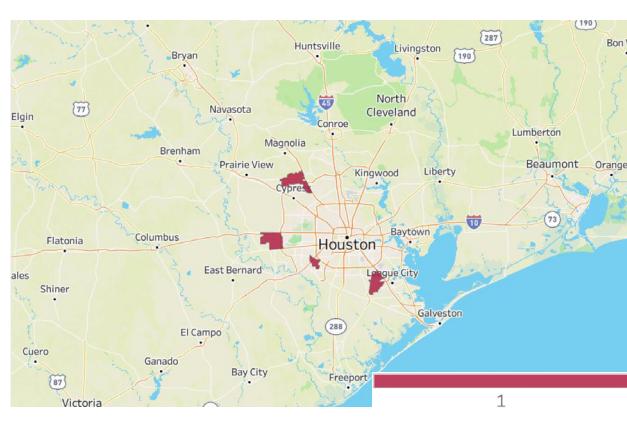
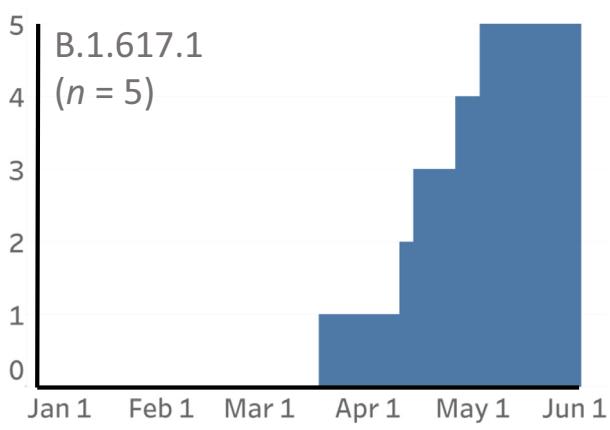
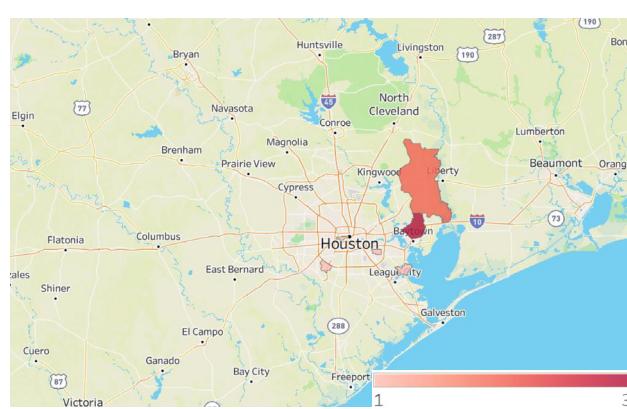
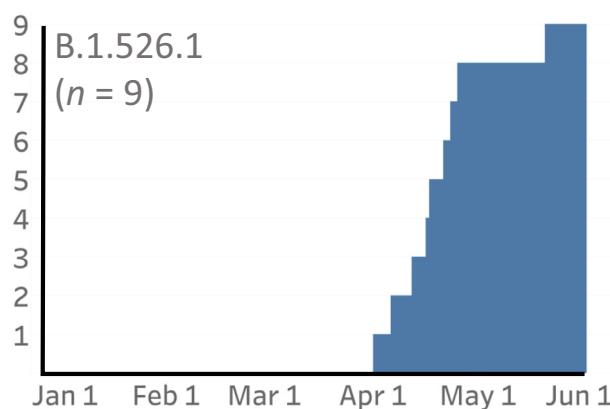
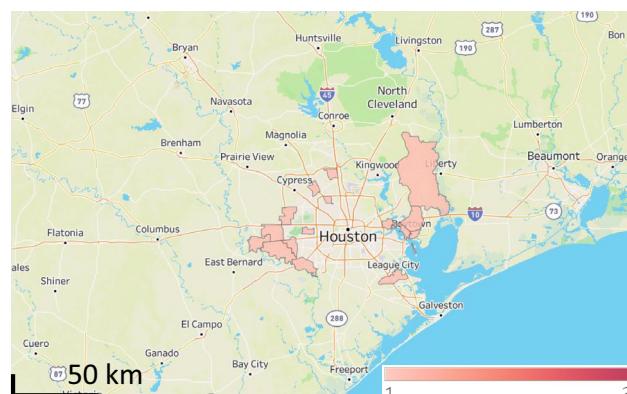
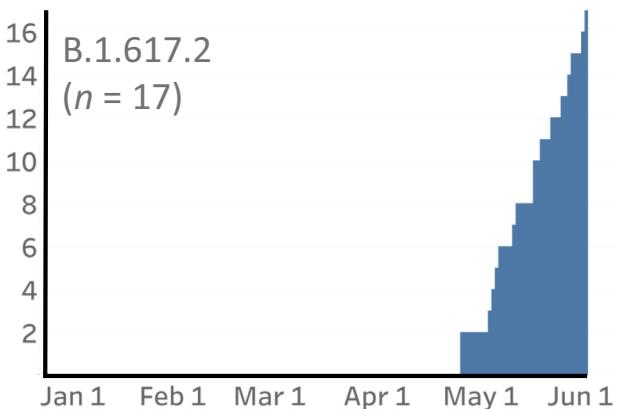


Figure 4 Cumulative increase in SARS-CoV-2 B.1.617.2, B.1.526.1, B.1.617.1, and B.1.351 variants and their distribution in metropolitan Houston, TX. The time frame used is January 1, 2021, through May 31, 2021. **Left column:** The cumulative increase in unique patients with each variant. **Right column:** The geospatial distribution of these variants based on the home address zip code for each patient. Figures were generated with Tableau version 2020.3.4.

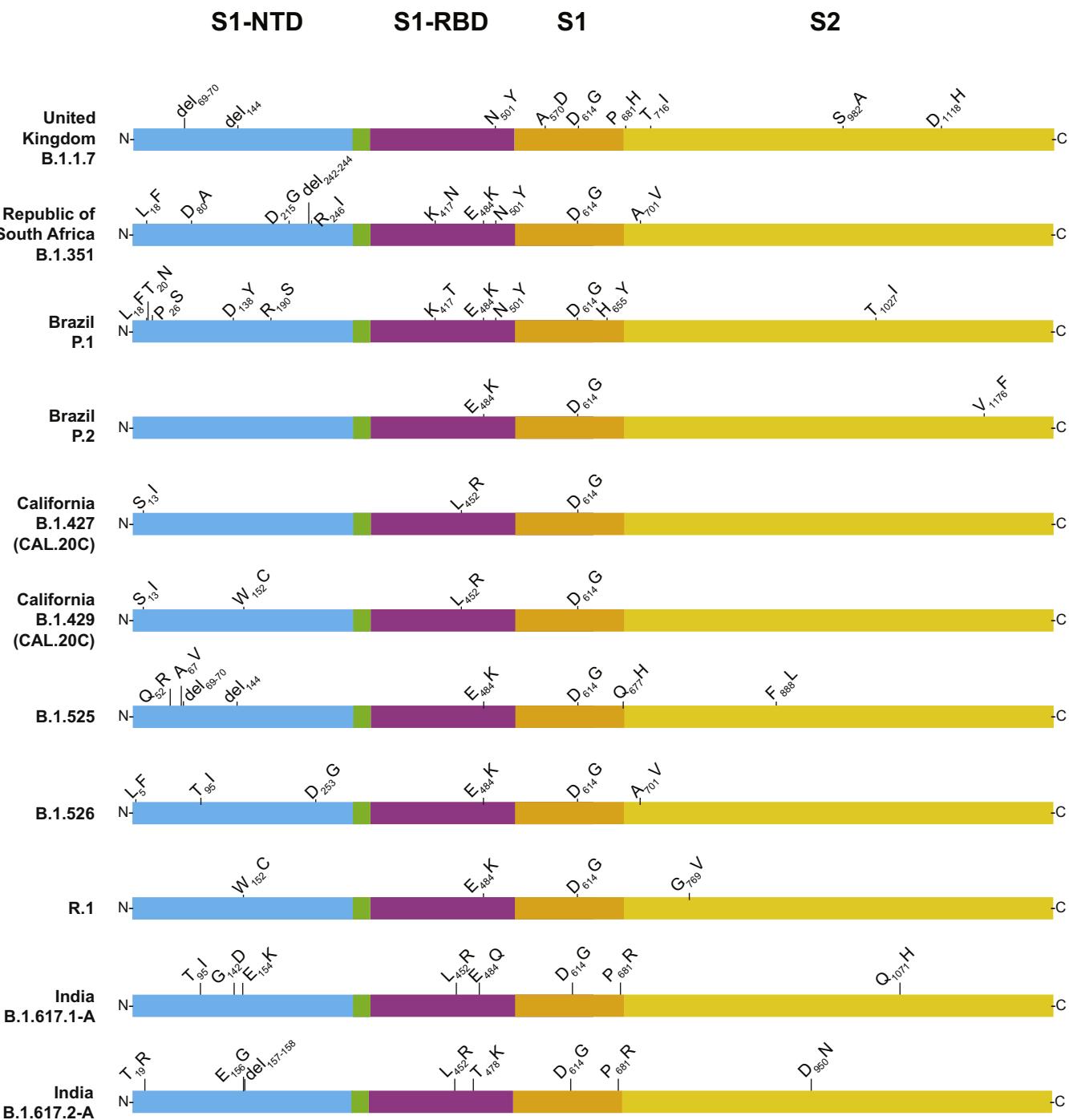


Figure 5 Structural changes present in spike protein of the major SARS-CoV-2 variants identified in the study, including variants of interest, variants of concern, and variant R.1. The figure is a modified version of one presented previously,¹⁰ with permission from Elsevier. NTD, amino-terminal domain; RBD, receptor-binding domain; S1, S1 domain; S2, S2 domain.

hypothesis that other VOCs and VOIs have significantly lower C_T values was tested. B.1.1.7 samples were removed for this analysis because their inclusion would confound the data. The data show that B.1.427/9 samples also had significantly lower C_T values; further analysis found that this signal was attributable to the results for the B.1.429 samples (Figure 7). C_T data for the P.2 and R.1 patient

samples were also significantly lower (Figure 7). Taken together, these observations are consistent with the idea that, on average, several common SARS-CoV-2 variants have significantly lower C_T values, a feature that may make them better able to disseminate and become dominant variants in the population. The sample sizes for the other VOCs and VOIs are not adequate for a meaningful analysis.

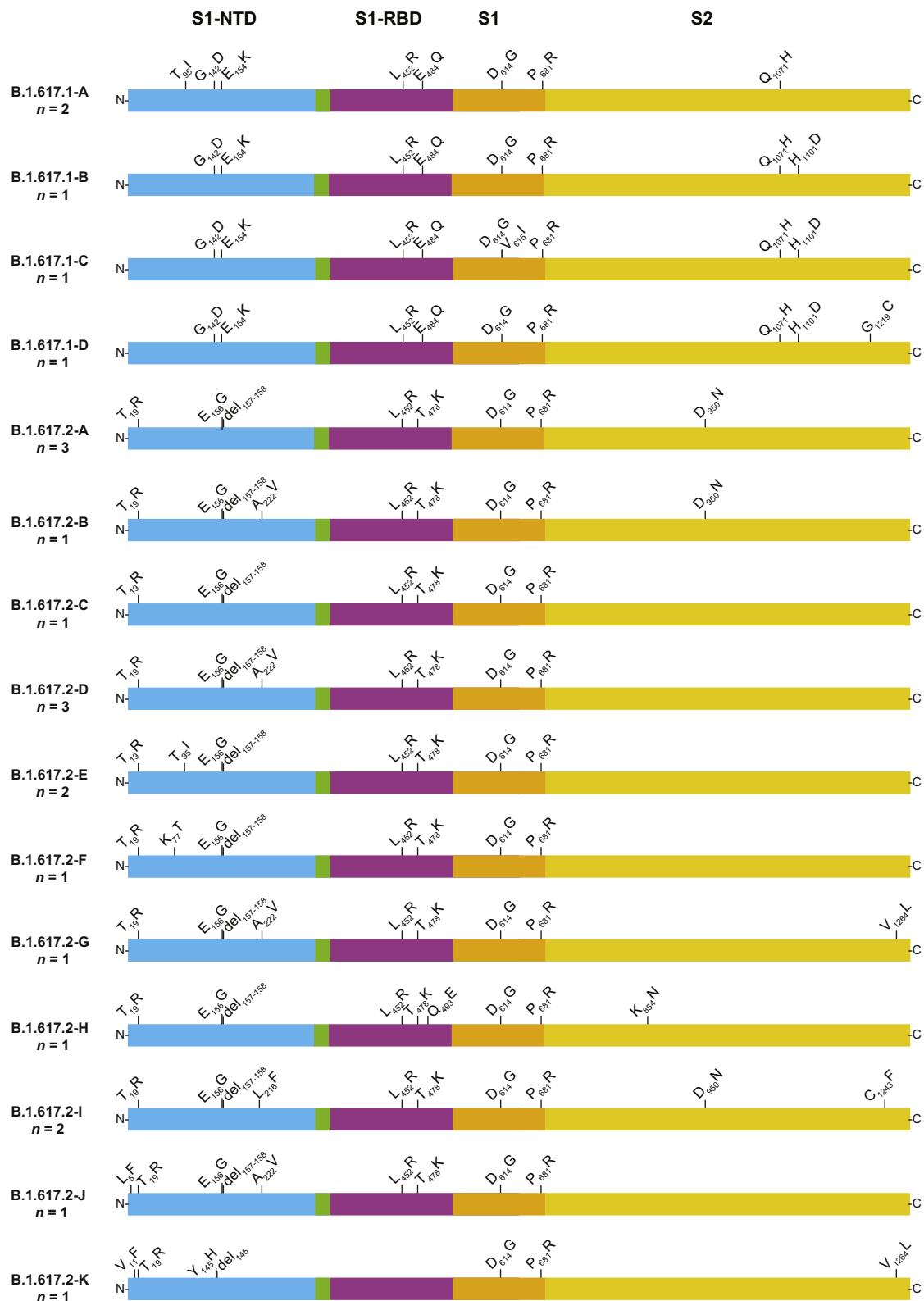


Figure 6 Structural changes present in spike protein of B.1.617-family variants. Four subvariants of B.1.617.1 and 10 subvariants of B.1.617.2 were identified. For the purpose of clarity, each subvariant was given an upper-case letter designation. The subvariants are listed (top to bottom) based on their decreasing abundance in The Global Initiative on Sharing Avian Influenza Data (GISAID) as of June 2, 2021. For example, B.1.617.1-A is the most common subvariant of B.1.617.1, and B.1.617.2-A is the most common subvariant of B.1.617.2. The number in parentheses below the subvariant designation denotes the number of patients with each subvariant identified in this study. Note that some annotation methods treat the E156G and del157-158 differently; we have used the GISAID annotation nomenclature. The figure is a modified version of one presented previously,¹⁰ with permission from Elsevier. NTD, amino-terminal domain; RBD, receptor-binding domain; S1, S1 domain; S2, S2 domain.

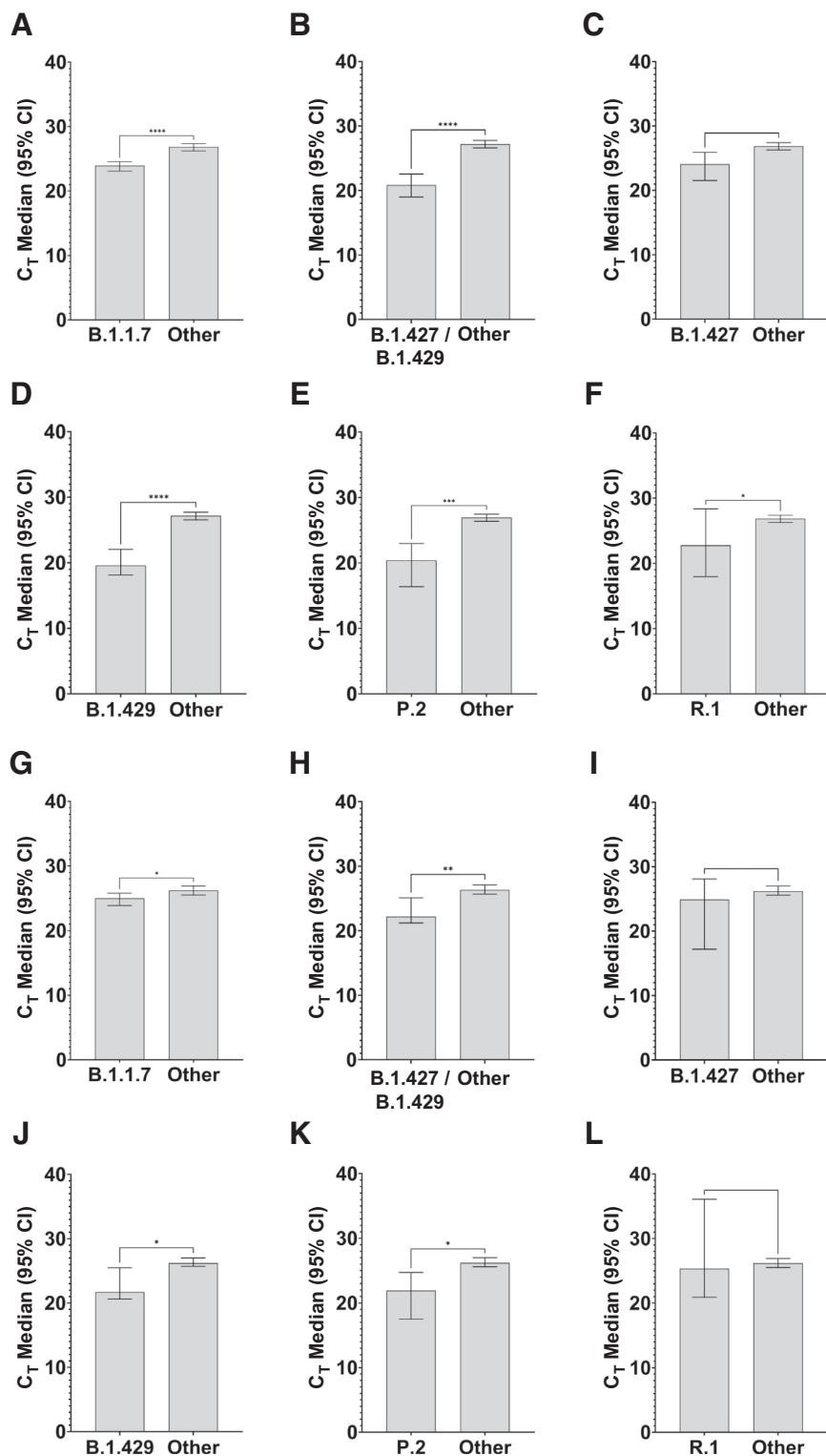


Figure 7 C_T value for every SARS-CoV-2 patient sample tested using the Abbott Alinity m (A–F) or Hologic Panther (G–L) assays, as described in *Materials and Methods*. Data are presented as median (95% CI) (A–L). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and **** $P < 0.0001$ (*U*-test).

Variant Geospatial Distribution

We next examined the geospatial distribution of all VOCs and VOIs in metropolitan Houston. With the exception of the B.1.351, B.1.526.1, B.1.617.1, and B.1.617.2 variants

(attributable to small sample sizes), patients infected with all other variants were dispersed broadly throughout metropolitan Houston, a finding consistent with the propensity of SARS-CoV-2 to spread rapidly between individuals (Figures 2, 3, and 4).

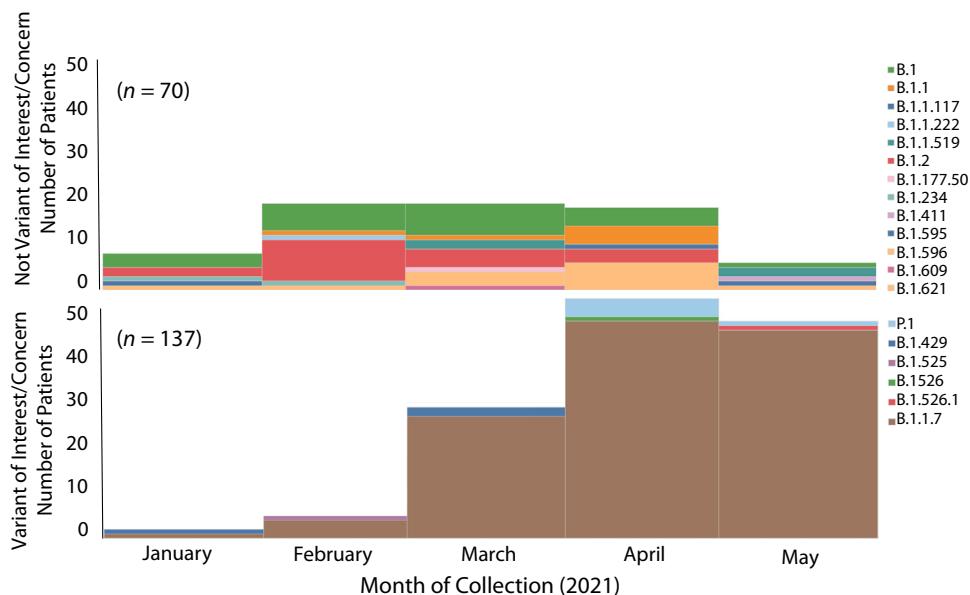


Figure 8 Number of COVID-19 vaccine breakthrough cases over time, by virus variant. The month of diagnosis and the infecting virus variant based on whole-genome sequencing are shown. A total of 207 vaccine breakthrough cases was identified in patients receiving either the Pfizer [$n = 181$ (87%)] or Moderna [$n = 26$ (13%)] vaccine.

E484 Spike Protein Amino Acid Changes and Convergent Evolution

Amino acid replacements at position E484 in spike protein have been of considerable research and public health interest, in part because they can decrease the efficacy of SARS-CoV-2 therapeutic antibodies and vaccine- or infection-induced adaptive immunity. Three hundred and sixty-three samples were identified with changes at E484 (E484K, $n = 353$; E484Q, $n = 9$; and E484D, $n = 1$) that occurred in many genetically diverse SARS-CoV-2 lineages were identified, some of which have not shared a recent common ancestor. For example, E484K polymorphism was found in samples from 69 patients infected with VOI P.2 and 43 patients infected with the newly described variant R.1.^{50–52} R.1 has the following core spike protein amino acid changes: W152L, E484K, D614G, and G769V (Outbreak.info, <https://outbreak.info/situation-reports?pango=r.1>, last accessed May 17, 2021). Some R.1 variants also contain R21T, L54F, S254P, or P116L changes. Of note, 11 patients infected with B.1.1.7 plus the E484K amino acid change, and one patient each infected with B.1.1.7 sample plus either an E484Q or an E484D amino acid change were identified. E484K replacement alters the immunologic profile of SARS-CoV-2,^{38,39,53–55} and Greaney et al⁵⁶ reported that E484Q reduced viral neutralization for some plasma samples.

N440K Spike Protein Replacement

The N440K amino acid change in spike protein has recently gained interest because samples with this

polymorphism have been reported to cause widespread COVID-19 in some states in India, increase viral titer *in vitro*, and have been associated with resistance to some candidate monoclonal antibody therapies.^{57,58} Eighteen patients with this N440K replacement were identified, and 10 patients had the identical combination of spike amino acid replacements: L18R, T95I, R158S, N440K, D614G, P681H, A688V, S735A, and T1027I. Two additional patients had SARS-CoV-2 of this same spike protein genotype with an additional T376I amino acid replacement. Pangolin categorized these strains as B.1. These 18 individuals were from 14 different zip codes dispersed throughout five counties in metropolitan Houston (data not shown). Of the 18 patients, 10 required hospitalization, and all were subsequently discharged.

Unexpected Identification of Samples with the B.1.1.7 Variant in Early December 2020

In work conducted contemporaneously with the present study, all genomes from earlier in the pandemic in Houston were routinely sequenced, including the uptick part of the third wave of disease occurring in November and December 2020 (Figure 1). Five patients in the first 10 days of December were identified with infections caused by B.1.1.7, an unexpected result because the first Houston Methodist patient previously documented with this VOC was identified in early January 2021,¹⁰ and the first Texas patient with B.1.1.7 was announced by state public health authorities on January 7, 2021 (Texas Department of State Health Services, <https://www.dshs.texas.gov/news/releases/2021/20210107a.aspx>, last accessed May 17, 2021). Thus, our genome data revise

these timelines. Based on genome sequences deposited in The Global Initiative on Sharing Avian Influenza Data (GISaid, <https://www.gisaid.org>, last accessed May 17, 2021), only five B.1.1.7 sequences from the United States were deposited with collection dates before these five Houston B.1.1.7 patients tested positive. Thus, these Houston patients are some of the earliest documented infections caused by the B.1.1.7 VOC in the United States, a finding that further highlights the importance of comprehensive genome sequencing of large populations from metropolitan areas with diverse patient populations.

Variants and Vaccine Breakthrough COVID-19 Cases

COVID-19 vaccines have remarkably high efficacy in preventing clinical infection caused by SARS-CoV-2, as shown by large randomized-controlled trials.^{59–61} Two mRNA-based vaccines given Emergency Use Authorization by the US Food and Drug Administration have been used widely in the United States (<https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine>, last accessed May 28, 2021; and <https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/moderna-covid-19-vaccine>, last accessed May 28, 2021), and extensively in the Houston Methodist health system. Despite the high efficacy of these two mRNA-based vaccines, a relatively small percentage of individuals who have received all recommended doses developed either asymptomatic or symptomatic SARS-CoV-2 infections.^{62–64} The contributors to vaccine breakthroughs are not fully understood, but there is concern that genetic variants of SARS-CoV-2 may play an outsized role, especially those with structural changes in spike protein that can alter immunologic characteristics. Consistent with this idea, McEwen et al⁶⁴ recently reported that all 20 vaccine breakthrough cases identified at the University of Washington were caused by VOCs. In contrast, a nationwide study found that 64% of breakthrough cases were caused by VOCs, although genome data were available from only 5% of reported cases included in the study.⁶³

To test the hypothesis that VOCs and VOIs were overrepresented among post-vaccination breakthrough infections among our COVID-19 cases, metadata available for the 12,476 patients reported herein was examined. Two hundred and twenty-four patients who met the criteria of vaccine breakthrough (ie, infection occurring >14 days after full vaccination was completed) were identified. SARS-CoV-2 genome sequence data were obtained for 207 cases (Table 1), of which, 72 (34.8%) required hospitalization. The 207 patients were infected with a heterogeneous array of variants, only some of which were VOIs or VOCs (Figure 8). In the aggregate, there was a significant increase in VOIs or VOCs among the breakthrough cases ($P < 0.001$) (Table 1). More importantly, the infecting viruses generally reflected the spectrum of SARS-CoV-2 variants circulating in the Houston

metropolitan region during the time of diagnosis of the vaccine breakthrough case. For example, in January and February 2021, many diverse variants were causing disease in the metroplex, and reflecting that fact, the viruses causing breakthrough cases were genetically heterogeneous (Figure 8). Similarly, as the B.1.1.7 VOC rose to prominence in Houston in March, April, and May, it caused the great majority of vaccine breakthrough cases (Figure 8). Of note, five breakthrough cases in April and May were caused by VOC P.1, a variant that was also increasing in disease frequency during this period (Figures 2, 3, and 4).

Discussion

The molecular population genomics of SARS-CoV-2 occurring in metropolitan Houston were analyzed with a focus on infections occurring early in 2021, from January 1 through May 31. This study was based on genome sequences from 12,476 ethnically, socioeconomically, and geographically diverse patients distributed throughout the metropolitan area. Infections caused by B.1.1.7 increased rapidly, and in the latter half of May caused 63% to 90% of all new cases in the population. Compared with non-B.1.1.7 patients, individuals infected with B.1.1.7 had significantly lower virus C_T values and a higher rate of hospitalization, but no difference in length of stay or mortality. Twenty two patients were infected with B.1.617-family variants, genotypes that are now causing extensive disease in India, elsewhere in Southeast Asia, and several areas of the United Kingdom.^{42–44}

A key finding from this study was the rapid growth trajectory of VOC B.1.1.7 in metropolitan Houston, an area with a population size of approximately 7 million. Patients infected with the B.1.1.7 VOC have significantly lower C_T values on initial diagnosis, but this has not been a universal finding.^{11,65–70} In the absence of quantitative virus cultures, the C_T value is viewed by many as a convenient proxy for virus load. C_T value in nasopharyngeal swabs taken from B.1.1.7 patients was significantly decreased compared with that in non-B.1.1.7 patients (Table 1 and Figure 7), a result consistent with previous reports.^{11,66,71–73} Thus, these data are consistent with the potential for enhanced transmissibility of B.1.1.7 because of higher nasopharyngeal virus loads. However, it is clear that there is no uniform relationship between C_T value and ability to disseminate. For example, patients infected with B.1.1.7 and high C_T values and non-B.1.1.7 patients with low C_T values were identified. Many factors contribute to SARS-CoV-2 transmission dynamics, including, but not limited to, behavioral characteristics of human populations, percentage of susceptible individuals, vaccination status, network structure, and biologic variation in the capacity of virus genotypes to survive and be successfully transmitted. Collectively, these findings stress the need for more information about the

relationship between C_T values, quantitative virus cultures, and specific genotypes of SARS-CoV-2.

A significantly increased hospitalization rate was identified for patients with B.1.1.7 compared with non-B.1.1.7 patients, but no significant difference was observed in the length of hospitalization or 28-day mortality (Table 1). Several studies have examined the relationship between disease severity and B.1.1.7.^{23–36} Patone et al³⁵ estimated the risk of critical care admission and overall mortality associated with B.1.1.7 compared with the original variant circulating in the United Kingdom among a large group of patients. They reported that patients infected with B.1.1.7 have significantly increased risk for critical care admission and mortality compared with patients not infected with B.1.1.7. However, the risk of mortality was linked to receiving critical care, not distinct virus genotype. They concluded that VOC B.1.1.7 causes more severe disease.

In the United Kingdom, at the end of February, the B.1.1.7 variant accounted for 98% of all COVID-19 cases (Wikipedia, https://en.wikipedia.org/wiki/lineage_b.1.1.7, last accessed June 9, 2021).²⁶ A similar rapid increase in B.1.1.7 has been reported in many countries, including Israel, France, Denmark, Norway, and Lebanon (Wikipedia, https://en.wikipedia.org/wiki/lineage_b.1.1.7, last accessed June 9, 2021). The current data show that this variant increased rapidly in metropolitan Houston since January 2021, and in the second half of May caused 63% to 90% of new COVID-19 cases daily. However, the increase in B.1.1.7 as percentage of new cases has occurred in the context of a substantial decrease in total COVID-19 cases in our metropolitan region (Figure 1). Although the precise cause of these seemingly disparate trends is unknown, we hypothesize that a relatively successful early vaccination campaign in the region coupled with heightened public awareness and concern about variants contributed to the decreasing case rate, whereas the increase in percentage of cases caused by B.1.1.7 is attributed to the capacity of this variant to transmit more rapidly than other variants. A contribution of a small but significant ability of B.1.1.7 to evade immunity induced by either natural infection or vaccination cannot be ruled out, and the current data are fully consistent with this idea (Table 1). In this regard, published data shows that B.1.1.7 differs in some immunologic characteristics compared with wild-type SARS-CoV-2.^{74–81}

SARS-CoV-2 variants with the E484K amino acid replacement are of particular concern in many areas, including Brazil, South Africa, and India (Center for Infectious Disease Research and Policy, <https://www.cidrap.umn.edu/news-perspective/2021/02/pfizer-moderna-vaccines-may-be-less-effective-against-b1351-variant>, last accessed May 17, 2021). Consistent with other studies, this study identified the E484K change in several genetically distinct lineages of the virus, a finding likely attributable to convergent evolution, as noted previously by others.^{38,39,53–55,82} In the United Kingdom, genome

sequencing efforts have identified the E484K change in some B.1.1.7 samples, although it remains a minor subpopulation (<https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers/variants-distribution-of-cases-data>, last accessed May 17, 2021).⁸³ The B.1.1.7 plus E484K variant has been reported infrequently elsewhere in the United States (Outbreak.info, <https://outbreak.info/situation-reports?pango=b.1.1.7&mut=s%3ae484k>, last accessed May 17, 2021). With the identification of this amino acid change in only 11 (0.3%) of the 3276 B.1.1.7 patients, the current data mirror the UK findings and other US findings.

The R.1 variant was first reported in Arizona in October 2020, and soon thereafter was identified in Canada and Japan (<https://outbreak.info/situation-reports?pango=r.1>, last accessed May 17, 2021).^{50–52} Cavanaugh et al⁵⁰ recently reported that an R.1 lineage variant was responsible for a COVID-19 outbreak in a skilled nursing facility in Kentucky in March 2021. The first Houston Methodist patient with the variant R.1 was identified in mid-December 2020, and its prevalence increased during the study period (Figures 2, 3, and 4). Interestingly, for unknown reasons, its prevalence plateaued by early April (Figures 2, 3, and 4).

Two hundred and seven symptomatic patients documented to be fully immunized and with a specimen taken for diagnosis >14 days after receiving their second dose of either the Pfizer [$n = 181$ patients (87%)] or Moderna [$n = 26$ patients (13%)] mRNA SARS-CoV-2 vaccine were studied. Because the virus genomes causing the vast majority of Houston Methodist Hospital cases (93%) were sequenced, these vaccine breakthrough cases were discovered to be caused by many diverse SARS-CoV-2 genotypes, including VOIs, VOCs, and many variants not assigned to either of these categories (Figure 8). The current results are similar to those reported by the CDC COVID-19 Vaccine Breakthrough Case Investigations Team,⁶³ but contrast with those reported recently by McEwen et al,⁶⁴ who found that all 20 breakthrough cases were caused by VOCs. Although the B.1.1.7 VOC caused most of the breakthrough cases from mid-April onward, this prominence reflected the rapid increase and abundance of this variant throughout the Houston area during that period. Clearly, more understanding is needed of the factors underlying vaccine breakthrough infections, and studies are underway to examine potential contributors.

Although extensive genomic data are not available, members of the B.1.617 variant family are contributing to the COVID-19 disease surge in India and other countries in Southeast Asia (World Health Organization, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—8-june-2021>, last accessed June 9, 2021).^{42–49} B.1.617 variants also have been documented to be increasing rapidly in many areas of the United Kingdom, and have been estimated to be 60% more transmissible than B.1.1.7, the variant that rapidly rose to dominate new infections in the United Kingdom.^{48,49,84} In this regard, the

identification of 22 patients in the Houston metropolitan area infected with variants B.1.617.1 and B.1.617.2 is concerning. One of the 22 Houston patients was diagnosed in mid-March 2021, which makes it one of the earliest documented cases of this variant in the United States, with only 11 cases identified before this, starting on February 25 (GISAID, <https://www.gisaid.org>, last accessed May 17, 2021). The relatively high number of B.1.617 subvariants ($n = 14$) was unexpected (Table 2 and Figure 6), and likely reflects the large population size of B.1.617-family variants worldwide. In this regard, several of our patients recently traveled outside the United States. The B.1.617-family variants have amino acid changes in spike protein (Figure 6) that have been linked to increased transmissibility and resistance to antibodies that are generated by natural infection or vaccination, and altered virulence in some studies.^{42–44,46,47} It will be important to continue to monitor SARS-CoV-2 genomes from patients in the Houston area to determine the rate of spread of the variants, and assess if new variants that arise have biomedically relevant phenotypes, such as enhanced virulence and immunologic escape.

Limitations

This study has several limitations. During the January 1 through May 31, 2021, study period, 269,341 cases of COVID-19 were reported in Harris County and its eight contiguous counties (USAfacts, https://usafactsstatic.blob.core.windows.net/public/data/covid-19/covid_confirmed_usafacts.csv, last accessed June 8, 2021). Thus, although 93% of all Houston Methodist cases identified during this period were sequenced, our genome sample represents only 4.6% of all reported cases in the metropolitan region. Our eight hospitals and outpatient clinics are geographically widely dispersed across the metropolitan region and serve patients who are demographically, socioeconomically, and geographically highly diverse. However, unless all SARS-CoV-2 genotypes are equally distributed throughout all populations in the Houston metropolitan region, our sample may underrepresent some SARS-CoV-2 genotypes causing COVID-19 in some populations, such as the homeless and other disenfranchised individuals. Our hospitals and clinics care mainly for adult patients, which means that SARS-CoV-2 variants causing pediatric cases are underrepresented in this study, although overall, the number of cases in this age group is relatively small. Finally, virtually all sequenced SARS-CoV-2 genomes were obtained from symptomatic patients, which may reflect the 48.4% overall hospitalization rate. Thus, our sample may underrepresent genotypes causing only asymptomatic carriage.

Summary

To summarize, in the latter half of May 2021, 63% to 90% of all new COVID-19 cases among ethnically, geographically, and socioeconomically diverse Houston Methodist

health care system patients were caused by the B.1.1.7 variant. Vaccine breakthrough cases were caused by SARS-CoV-2 infections that are genetically diverse and largely reflect the genotypes that are circulating and abundant in the community. Identification of 22 patients with the B.1.617 family of variants and 11 patients with B.1.1.7 plus E484K in metropolitan Houston is cause for heightened concern. Although our sample represents only 4.6% of all reported COVID-19 cases in the Houston area, it is reasonable to extrapolate that B.1.617-family variants have caused ≈ 400 cases in our region. The rate and extent of spread of these variants should be monitored closely by rapid genome sequencing, coupled with analysis of patient metadata, including disease severity and mortality. This is an especially pressing issue for B.1.617-family variants because B.1.617.2 has become abundant and is outcompeting B.1.1.7 in many areas of the United Kingdom (*Outbreak.info*, <https://outbreak.info/location-reports?loc=gbr>, last accessed June 8, 2021; and The Telegraph, <https://www.telegraph.co.uk/global-health/science-and-disease/in-dian-variant-covid-coronavirus-uk>, last accessed June 8, 2021). Moreover, these data show a high rate of hospitalization for patients infected with B.1.617 variants (Table 2), a finding consistent with data recently reported by Public Health England (GOV.UK, <https://www.gov.uk/government/news/confirmed-cases-of-covid-19-variants-identified-in-uk>, last accessed June 9, 2021).

Acknowledgments

We thank the many molecular technologists and volunteers in the Molecular Diagnostics Laboratory and Methodist Research Institute for their efforts; Dr. Marc Boom and Dr. Dirk Sostman for support; Houston philanthropists for support to the Houston Methodist Academic Institute Infectious Diseases Fund; Jessica W. Podnar and personnel in the University of Texas Genome Sequencing and Analysis Facility for sequencing some of the genomes in this study; Trina Trinh, Hung-Che Kuo, and G. Nguyen for genome sequencing support; the originating and submitting laboratories of the SARS-CoV-2 genome sequences from The Global Initiative on Sharing Avian Influenza Data (GISAID) EpiFlu Database used in some of the work presented herein; many colleagues for critical reading of the manuscript and suggesting improvements; and Dr. Sasha Pejerrey, Dr. Kathryn Stockbauer, Adrienne Winston, and Dr. Heather McConnell for help with figures, tables, and editorial contributions.

Author Contributions

J.M.M. conceptualized and designed the project; R.J.O., P.A.C., S.W.L., S.S., R.O., M.N., J.J.D., P.Y., M.O.S., L.P., K.R., M.N.S., R.G., J.C., R.M.T., A.B., I.J.F., and J.G. performed research; all authors wrote the manuscript.

References

- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020, 395:497–506
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus I, Research T: A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020, 382:727–733
- Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsui HW, Lo SK, Chan KH, Poon VK, Chan WM, Ip JD, Cai JP, Cheng VC, Chen H, Hui CK, Yuen KY: A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020, 395:514–523
- Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ: A new coronavirus associated with human respiratory disease in China. *Nature* 2020, 579:265–269
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses: The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020, 5:536–544
- Wang C, Horby PW, Hayden FG, Gao GF: A novel coronavirus outbreak of global health concern. *Lancet* 2020, 395:470–473
- Allel K, Tapia-Munoz T, Morris W: Country-level factors associated with the early spread of COVID-19 cases at 5, 10 and 15 days since the onset. *Glob Public Health* 2020, 15:1589–1602
- Long SW, Olsen RJ, Christensen PA, Bernard DW, Davis JR, Shukla M, Nguyen M, Ojeda Saavedra M, Cantu CC, Yerramilli P, Pruitt L, Subedi S, Hendrickson H, Eskandari G, Kumaraswami M, McLellan JS, Musser JM: Molecular architecture of early dissemination and evolution of the SARS-CoV-2 virus in metropolitan Houston, Texas. *bioRxiv* 2020, 11. e02707-20
- Long SW, Olsen RJ, Christensen PA, Bernard DW, Davis JJ, Shukla M, Nguyen M, Saavedra MO, Yerramilli P, Pruitt L, Subedi S, Kuo HC, Hendrickson H, Eskandari G, Nguyen HAT, Long JH, Kumaraswami M, Goike J, Boutz D, Gollihar J, McLellan JS, Chou CW, Javanmardi K, Finkelstein II, Musser JM: Molecular architecture of early dissemination and massive second wave of the SARS-CoV-2 virus in a major metropolitan area. *MBio* 2020, 11:e02707–e02720
- Long SW, Olsen RJ, Christensen PA, Subedi S, Olson R, Davis JJ, Saavedra MO, Yerramilli P, Pruitt L, Reppond K, Shyer MN, Cambric J, Finkelstein II, Gollihar J, Musser JM: Sequence analysis of 20,453 severe acute respiratory syndrome coronavirus 2 genomes from the Houston metropolitan area identifies the emergence and widespread distribution of multiple isolates of all major variants of concern. *Am J Pathol* 2021, 191:983–992
- Musser JM, Olsen RJ, Christensen PA, Long SW, Subedi S, Davis JJ, Gollihar J: Rapid, widespread, and preferential increase of SARS-CoV-2 B.1.1.7 variant in Houston, TX, revealed by 8,857 genome sequences. *medRxiv* 2021. [Preprint] doi:10.1101/2021.03.16.21253753
- Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis JJ, Brettin T, Xia F, Musser JM: Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* isolates, Houston, Texas: unexpected abundance of clonal group 307. *MBio* 2017, 8:e00489-17
- Long SW, Beres SB, Olsen RJ, Musser JM: Absence of patient-to-patient intrahospital transmission of *Staphylococcus aureus* as determined by whole-genome sequencing. *MBio* 2014, 5:e01692-14
- Wright AM, Beres SB, Consamus EN, Long SW, Flores AR, Barrios R, Richter GS, Oh SY, Garufi G, Maier H, Drews AL, Stockbauer KE, Cernoch P, Schneewind O, Olsen RJ, Musser JM: Rapidly progressive, fatal, inhalation anthrax-like infection in a human: case report, pathogen genome sequencing, pathology, and coordinated response. *Arch Pathol Lab Med* 2011, 135:1447–1459
- Nasser W, Beres SB, Olsen RJ, Dean MA, Rice KA, Long SW, Kristinsson KG, Gottfredsson M, Vuopio J, Raisanen K, Caugant DA, Steinbakk M, Low DE, McGeer A, Darenberg J, Henriques-Normark B, Van Beneden CA, Hoffmann S, Musser JM: Evolutionary pathway to increased virulence and epidemic group A *Streptococcus* disease derived from 3,615 genome sequences. *Proc Natl Acad Sci U S A* 2014, 111:E1768–E1776
- Kachroo P, Eraso JM, Beres SB, Olsen RJ, Zhu L, Nasser W, Bernard PE, Cantu CC, Saavedra MO, Arredondo MJ, Strope B, Do H, Kumaraswami M, Vuopio J, Grondahl-Yli-Hannuksela K, Kristinsson KG, Gottfredsson M, Pesonen M, Pensar J, Davenport ER, Clark AG, Corander J, Caugant DA, Gaini S, Magnussen MD, Kubiak SL, Nguyen HAT, Long SW, Porter AR, DeLeo FR, Musser JM: Integrated analysis of population genomics, transcriptomics and virulence provides novel insights into *Streptococcus pyogenes* pathogenesis. *Nat Genet* 2019, 51:548–559
- Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI, Sheffield C-GG, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC: Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020, 182:812–827.e19
- Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR, Mirchandani D, Scharton D, Bilello JP, Ku Z, An Z, Kalveram B, Freiberg AN, Menachery VD, Xie X, Plante KS, Weaver SC, Shi PY: Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 2021, 592:116–121
- Daniloski Z, Jordan TX, Ilmain JK, Guo X, Bhabha G, tenOever BR, Sanjana NE: The spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *Elife* 2021, 10:e65365
- Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD, Rangarajan ES, Pan A, Vanderheiden A, Suthar MS, Li W, Izard T, Rader C, Farzan M, Choe H: SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun* 2020, 11:6013
- Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole A, Southgate J, Johnson R, Jackson B, Nascimento FF, Rey SM, Nicholls SM, Colquhoun RM, da Silva Filipe A, Shepherd J, Pascall DJ, Shah R, Jesudason N, Li K, Jarrett R, Pacchiarini N, Bull M, Geidelberg L, Siveroni I, Consortium C-U, Goodfellow I, Loman NJ, Pybus OG, Robertson DL, Thomson EC, Rambaut A, Connor TR: Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* 2021, 184:64–75.e11
- Benton DJ, Wrobel AG, Roustan C, Borg A, Xu P, Martin SR, Rosenthal PB, Skehel JJ, Gamblin SJ: The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. *Proc Natl Acad Sci U S A* 2021, 118:e0222586118
- Grubaugh ND, Hodcroft EB, Fauver JR, Phelan AL, Cevik M: Public health actions to control new SARS-CoV-2 variants. *Cell* 2021, 184: 1127–1132
- Walensky RP, Walke HT, Fauci AS: SARS-CoV-2 variants of concern in the United States—challenges and opportunities. *JAMA* 2021, 325:1037–1038
- Mascola JR, Graham BS, Fauci AS: SARS-CoV-2 viral variants—tackling a moving target. *JAMA* 2021, 325:1261–1262
- Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, Wenseleers T, Gimma A, Waites W, Wong KLM, van Zandvoort K, Silverman JD, Group CC-W, Consortium C-GU, Diaz-Ordaz K, Keogh R, Eggo RM, Funk S, Jit M, Atkins KE, Edmunds WJ:

- Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 2021; 372:eabg3055
27. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, Hinsley WR, Laydon DJ, Dabrerera G, O'Toole Á, Amato R, Ragonnet-Cronin M, Harrison I, Jackson B, Ariani CV, Boyd O, Loman NJ, McCrone JT, Gonçalves S, Jorgensen D, Myers R, Hill V, Jackson DK, Gaythorpe K, Groves N, Sillitoe J, Kwiatkowski DP, Flaxman S, Ratmann O, Bhatt S, Hopkins S, Gandy A, Rambaut A, Ferguson NM: Transmission of SARS-CoV-2 lineage B.1.1.7 in England: insights from linking epidemiological and genetic data. *medRxiv* 2021. [Preprint] doi:[10.1101/2020.12.30.20249034](https://doi.org/10.1101/2020.12.30.20249034)
 28. Graham MS, Sudre CH, May A, Antonelli M, Murray B, Varsavsky T, Kläser K, Canas LS, Molteni E, Modat M, Drew DA, Nguyen LH, Polidori L, Selvachandran S, Hu C, Capdevila J, Hammers A, Chan AT, Wolf J, Spector TD, Steves CJ, Ourself S: Changes in symptomatology, re-infection and transmissibility associated with SARS-CoV-2 variant B.1.1.7: an ecological study. *Lancet Public Health* 2021; 6:e335–e345
 29. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L: Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ* 2021; 372:n579
 30. Lumley SF, Rodger G, Constantinides B, Sanderson N, Chau KK, Street TL, et al: An observational cohort study on the incidence of SARS-CoV-2 infection and B.1.1.7 variant infection in healthcare workers by antibody and vaccination status. *Clin Infect Dis* 2021; ciab608
 31. Washington NL, Gangavarapu K, Zeller M, Bolze A, Cirulli ET, Schiabor Barrett KM, et al: Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell* 2021; 184: 2587–2594.e7
 32. Kissler SM, Fauver JR, Mack C, Tai CG, Breban MI, Watkins AE, Samant RM, Anderson DJ, Ho DD, Grubaugh ND, Grad YH: Densely sampled viral trajectories suggest longer duration of acute infection with B.1.1.7 variant relative to non-B.1.1.7 SARS-CoV-2. *medRxiv* [Preprint]. doi:[10.1101/2021.02.16.21251535](https://doi.org/10.1101/2021.02.16.21251535)
 33. Galloway SE, Paul P, MacCannell DR, Johansson MA, Brooks JT, MacNeil A, Slayton RB, Tong S, Silk BJ, Armstrong GL, Biggerstaff M, Dugan VG: Emergence of SARS-CoV-2 B.1.1.7 lineage - United States, December 29, 2020–January 12, 2021. *MMWR Morb Mortal Wkly Rep* 2021; 70:95–99
 34. Alpert T, Lasek-Nesselquist E, Brito AF, Valesano AL, Rothman J, MacKay MJ, et al: Early introductions and community transmission of SARS-CoV-2 variant B.1.1.7 in the United States. *medRxiv* 2021. [Preprint] doi:[10.1101/2021.02.10.21251540](https://doi.org/10.1101/2021.02.10.21251540)
 35. Patone M, Thomas K, Hatch R, Tan PS, Coupland C, Liao W, Mouncey P, Harrison D, Rowan K, Horby P, Watkinson P, Hippisley-Cox J: Analysis of severe outcomes associated with the SARS-CoV-2 variant of concern 202012/01 in England using ICNARC Case Mix Programme and QResearch databases. *medRxiv* 2021, [Preprint] doi:[10.1101/2021.03.11.21253364](https://doi.org/10.1101/2021.03.11.21253364)
 36. Piantham C, Linton NM, Nishiura H, Ito K: Estimating the increased transmissibility of the B.1.1.7 strain over previously circulating strains in England using frequencies of GISAID sequences and the distribution of serial intervals. *medRxiv* 2021. [Preprint] doi:[10.1101/2021.03.17.21253775](https://doi.org/10.1101/2021.03.17.21253775)
 37. Naveca F, Nascimento V, Souza V, Corado A, Nascimento F, Silva G, Costa Á, Duarte D, Pessoa K, Gonçalves L, Brandão MJ, Jesus M, Fernandes C, Pinto R, Silva M, Mattos T, Wallau GL, Siqueira MM, Resende PC, Delatorre E, Gräf T, Bello G: Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the spike protein. *virological.org* 2021
 38. Voloch CM, Silva FRD, de Almeida LGP, Cardoso CC, Brustolini OJ, Gerber AL, Guimarães APdC, Mariani D, Costa RMd, Ferreira OC, Cavalcanti AC, Frauches TS, de Mello CMB, Galliez RM, Faffe DS, Castilheiros TMPP, Tanuri A, de Vasconcelos ATR: Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *J Virol* 2021; 95:e00119-21
 39. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al: Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* 2020. [Preprint] doi:[10.1101/2020.12.21.20248640](https://doi.org/10.1101/2020.12.21.20248640)
 40. Bobeck SM, Janowska K, McDowell S, Mansouri K, Parks R, Stalls V, Kopp M, Manne K, Li D, Wiehe K, Saunders K, Edwards R, Korber B, Haynes B, Henderson R, Acharya P: Effect of natural mutations of SARS-CoV-2 on spike structure, conformation and antigenicity. *Science* 2021; 373:ea6226
 41. Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E: Emergence of a novel SARS-CoV-2 variant in southern California. *JAMA* 2021; 325:1324–1326
 42. Cherian S, Potdar V, Jadhav S, Yadav P, Gupta N, Das M, Rakshit P, Singh S, Abraham P, Panda S: Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. *bioRxiv* 2021. [Preprint] doi:[10.1101/2021.04.22.440932](https://doi.org/10.1101/2021.04.22.440932)
 43. Yadav PD, Sapkal GN, Abraham P, Ella R, Deshpande G, Patil DY, Nyayanit DA, Gupta N, Sahay RR, Shete AM, Panda S, Bhargava B, Mohan VK: Neutralization of variant under investigation B.1.617 with sera of BBV152 vaccinees. *Clin Infect Dis* 2021; ciab411
 44. Yadav PD, Mohandas S, Shete AM, Nyayanit DA, Gupta N, Patil DY, Sapkal GN, Potdar V, Kadam M, Kumar A, Kumar S, Suryavanshi D, Mote CS, Abraham P, Panda S, Bhargava B: SARS CoV-2 variant B.1.617.1 is highly pathogenic in hamsters than B.1 variant. *bioRxiv* 2021. [Preprint] doi:[10.1101/2021.05.05.442760](https://doi.org/10.1101/2021.05.05.442760)
 45. Hoffmann M, Hofmann-Winkler H, Krüger N, Kempf A, Nehlmeier I, Graichen L, Sidarovich A, Moldenhauer A-S, Winkler MS, Schulz S, Jäck H-M, Stankov MV, Behrens GMN, Pöhlmann S: SARS-CoV-2 variant B.1.617 is resistant to bamlanivimab and evades antibodies induced by infection and vaccination. *Cell Rep* 2021; 36:109415
 46. Ferreira I, Datir R, Kemp S, Papa G, Rakshit P, Singh S, et al: The Indian SARS-CoV-2 Genomics Consortium (INSACOG); The CIT-IID-NIHR BioResource COVID-19 Collaboration; The Genotype to Phenotype Japan (G2P-Japan) Consortium: SARS-CoV-2 B.1.617.2 Delta variant replication, sensitivity to neutralising antibodies and vaccine breakthrough. *bioRxiv* 2021. [Preprint] doi:[10.1101/2021.05.08.443253](https://doi.org/10.1101/2021.05.08.443253)
 47. Deng X, Garcia-Knight MA, Khalid MM, Servellita V, Wang C, Morris MK, et al: Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. *Cell* 2021; 184:3426–3437.e8
 48. Challen R, Dyson L, Overton CE, Guzman-Rincon LM, Hill EM, Stage HB, Brooks-Pollock E, Pellis L, Scarabel F, Pascall DJ, Blomquist P, Tildesley M, Williamson D, Siegert S, Xiong X, Youngman B, Read JM, Gog JR, Keeling MJ, Danon L: Early epidemiological signatures of novel SARS-CoV-2 variants: establishment of B.1.617.2 in England. *medRxiv* 2021. [Preprint] doi:[10.1101/2021.06.05.21258365](https://doi.org/10.1101/2021.06.05.21258365)
 49. Vöhringer HS, Sanderson T, Sinnott M, De Maio N, Nguyen T, Goater R, Schwach F, Harrison I, Hellewell J, Ariani C, Gonçalves S, Jackson D, Johnston I, Jung AW, Saint C, Sillitoe J, Suciu M, Goldman N, Panovska-Griffiths J; The Wellcome Sanger Institute Covid-19 Surveillance Team; The COVID-19 Genomics UK Consortium; Birney E, Volz E, Funk S, Kwiatkowski D, Chand M, Martincorena I, Barrett JC, Gerstung M: Genomic reconstruction of the SARS-CoV-2 epidemic across England from September 2020 to May 2021. *medRxiv* 2021. [Preprint] doi:[10.1101/2021.05.22.21257633](https://doi.org/10.1101/2021.05.22.21257633)
 50. Cavanaugh AM, Fortier S, Lewis P, Arora V, Johnson M, George K, Tobias J, Lunn S, Miller T, Thoroughman D, Spicer KB: COVID-19 outbreak associated with a SARS-CoV-2 R.1 lineage variant in a skilled nursing facility after vaccination program - Kentucky, March 2021. *MMWR Morb Mortal Wkly Rep* 2021; 70:639–643

51. Hirotsu Y, Omata M: Detection of R.1 lineage severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with spike protein W152L/E484K/G769V mutations in Japan. *PLOS Pathogens* [Epub] 2021. doi:[10.1371/journal.ppat.1009619](https://doi.org/10.1371/journal.ppat.1009619)
52. Tokumasu R, Weeraratne D, Snowdon J, Parida L, Kudo M, Koyama T: Introductions and evolutions of SARS-CoV-2 strains in Japan. *medRxiv* 2021. [Preprint] doi:[10.1101/2021.02.26.21252555](https://doi.org/10.1101/2021.02.26.21252555)
53. Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM, Theel ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH, Diamond MS, Whelan SPJ: Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* 2021, 29:477–488.e4
54. Liu H, Wei P, Zhang Q, Chen Z, Aviszus K, Downing W, Peterson S, Reynoso L, Downey GP, Frankel SK, Kappler J, Marrack P, Zhang G: 501Y.V2 and 501Y.V3 variants of SARS-CoV-2 lose binding to bamlanivimab in vitro. *MAbs* 2021, 13:1919285
55. Zhou D, Dejnirattisai W, Supasa P, Liu C, Mentzer AJ, Ginn HM, et al: Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* 2021, 184:2348–2361.e6
56. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD: Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 2021, 29: 463–476.e6
57. Tandell D, Gupta D, Sah V, Harinivas Harshan K: N440K variant of SARS-CoV-2 has higher infectious fitness. *bioRxiv* 2021. [Preprint] doi:[10.1101/2021.04.30.441434](https://doi.org/10.1101/2021.04.30.441434)
58. Rani PR, Imran M, Lakshmi JV, Jolly B, Jain A, Surekha A, Senthivel V, Chandrasekhar P, Divakar MK, Srinivasulu D, Bhoyer RC, Vanaja PR, Scaria V, Sivasubbu S: Symptomatic reinfection of SARS-CoV-2 with spike protein variant N440K associated with immune escape. *J Med Virol* 2021, 93:4163–4165
59. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Perez Marc G, Moreira ED, Zerbini C, Bailey R, Swanson KA, Roychoudhury S, Koury K, Li P, Kalina WV, Cooper D, French RW Jr, Hammitt LL, Tureci O, Nell H, Schaefer A, Unal S, Tresnan DB, Mather S, Dormitzer PR, Sahin U, Jansen KU, Gruber WC; C4591001 Clinical Trial Group: Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med* 2020, 383:2603–2615
60. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al: Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med* 2021, 384:403–416
61. Sadoff J, Gray G, Vandebosch A, Cárdenas V, Shukarev G, Grinsztejn B, Goepfert PA, Truyers C, Fennema H, Spiessens B, Offergeld K, Schepers G, Taylor KL, Robb ML, Treanor J, Barouch DH, Stoddard J, Ryser MF, Marovich MA, Neuzil KM, Corey L, Cauwenberghs N, Tanner T, Hardi K, Ruiz-Guiñazú J, Le Gars M, Schuitemaker H, Van Hoof J, Struyf F, Douoguih M: Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *N Engl J Med* 2021, 384:2187–2201
62. Hacisuleyman E, Hale C, Saito Y, Blachere NE, Bergh M, Conlon EG, Schaefer-Babajew DJ, DaSilva J, Muecksch F, Gaebler C, Lifton R, Nussenzweig MC, Hatziloannou T, Bieniasz PD, Darnell RB: Vaccine breakthrough infections with SARS-CoV-2 variants. *N Engl J Med* 2021, 384:2212–2218
63. CDC COVID-19 Vaccine Breakthrough Case Investigations Team: COVID-19 vaccine breakthrough infections reported to CDC - United States, January 1-April 30, 2021. *MMWR Morb Mortal Wkly Rep* 2021, 70:792–793
64. McEwen AE, Cohen S, Bryson-Cahn C, Liu C, Pergam SA, Lynch J, Schippers A, Strand K, Whimby E, Mani NS, Zelikoff AJ, Makarewicz VA, Brown ER, Mohamed Bakhsh SA, Baker NR, Castor J, Livingston RJ, Huang M-L, Jerome KR, Greninger AL, Roychoudhury P: Variants of concern are overrepresented among post-vaccination breakthrough infections of SARS-CoV-2 in Washington State. *Clin Infect Dis* 2021:ciab581
65. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Megraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S: Traces of human migrations in Helicobacter pylori populations. *Science* 2003, 299:1582–1585
66. Golubchik T, Lythgoe KA, Hall M, Ferretti L, Fryer HR, MacIntyre-Cockett G, de Cesare M, Trebes A, Piazza P, Buck D, Todd JA, Fraser C, Bonsall D: Early analysis of a potential link between viral load and the N501Y mutation in the SARS-CoV-2 spike protein. *medRxiv* 2021. [Preprint] doi:[10.1101/2021.01.12.20249080](https://doi.org/10.1101/2021.01.12.20249080)
67. Walker AS, Vihta K-D, Gethings O, Pritchard E, Jones J, House T, Bell I, Bell JI, Newton JN, Farrar J, Diamond I, Studley R, Rourke E, Hay J, Hopkins S, Crook D, Peto T, Matthews PC, Eyre DW, Stoesser N, Pouwels KB: Increased infections, but not viral burden, with a new SARS-CoV-2 variant. *medRxiv* 2021. [Preprint] doi:[10.1101/2021.01.13.21249721](https://doi.org/10.1101/2021.01.13.21249721)
68. Alizon S, Selinger C, Sofonea MT, Haim-Boukobza S, Giannoli J-M, Ninove L, Pillet S, Vincent T, de Rougemont A, Tumiotto C, Solis M, Stephan R, Bressollette-Bodin C, Salmona M, L'Honneur A-S, Behillil S, Lefevre C, Dina J, Hantz S, Hartard C, Veyer D, Delagrèverie HM, Fourati S, Visseaux B, Henquell C, Lina B, Foullonge V, Burrel S: Epidemiological and clinical insights from SARS-CoV-2 RT-PCR cycle amplification values. *medRxiv* 2021. [Preprint] doi:[10.1101/2021.03.15.21253653](https://doi.org/10.1101/2021.03.15.21253653)
69. Hay JA, Kennedy-Shaffer L, Kanjilal S, Lennon NJ, Gabriel SB, Lipsitch M, Mina MJ: Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science* 2021, 373:eab0635
70. Ratcliff J, Nguyen D, Fish M, Rhynne J, Jennings A, Williams S, Al-Beidh F, Bonsall D, Evans A, Golubchik T, Gordon AC, Lamikanra A, Tsang P, Ciccone N, Leuscher U, Slack W, Laing E, Mouncey PR, Ziyangeta S, Olivera M, Ploeg R, Rowan KM, Shankar-Hari M, Roberts DJ, Menon DK, Estcourt L, Simmonds P, Harvala H: Virological and serological characterization of critically ill patients with COVID-19 in the UK: a special focus on variant detection. *J Infect Dis* 2021;jiab283
71. Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, Scott R, Sconza R, Price J, Margaritis M, Bergstrom M, Spyri MJ, Miralhes PB, Grant P, Kirk S, Valerio C, Mangera Z, Prabhahar T, Moreno-Cuesta J, Arulkumaran N, Singer M, Shin GY, Sanchez E, Paraskevopoulou SM, Pillay D, McKendry RA, Mirfenderesky M, Houlihan CF, Nastouli E: Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *Lancet Infect Dis* 2021, 21:1246–1256
72. Lyngse FP, Mølbak K, Skov RL, Christiansen LE, Mortensen LH, Albertsen M, Møller CH, Krause TG, Rasmussen M, Michaelsen TY, Voldstedlund M, Fonager J, Steenhard N, Kirkeby C: Increased transmissibility of SARS-CoV-2 lineage B.1.1.7 by age and viral load: evidence from Danish households. *medRxiv* 2021. doi: [10.1101/2021.04.16.21255459](https://doi.org/10.1101/2021.04.16.21255459)
73. Kidd M, Richter A, Best A, Cumley N, Mirza J, Percival B, Mayhew M, Mogram O, Ashford F, White T, Moles-Garcia E, Crawford L, Bosworth A, Atabani SF, Plant T, McNally A: S-variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral load in samples tested by TaqPath polymerase chain reaction. *J Infect Dis* 2021, 223:1666–1670
74. Supasa P, Zhou D, Dejnirattisai W, Liu C, Mentzer AJ, Ginn HM, et al: Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell* 2021, 184:2201–2211.e7
75. Emery KRW, Golubchik T, Aley PK, Ariani CV, Angus B, Bibi S, et al: Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): an exploratory analysis of a randomised controlled trial. *Lancet* 2021, 397: 1351–1362
76. Collier DA, De Marco A, Ferreira I, Meng B, Datir RP, Walls AC, et al: Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* 2021, 593:136–141

-
- 77. Muik A, Wallisch AK, Sanger B, Swanson KA, Muhl J, Chen W, Cai H, Maurus D, Sarkar R, Tureci O, Dormitzer PR, Sahin U: Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. *Science* 2021, 371:1152–1153
 - 78. Mahase E: Covid-19: where are we on vaccines and variants? *BMJ* 2021, 372:n597
 - 79. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD, Graham BS, Mascola JR, Chang JY, Yin MT, Sobieszczuk M, Kyratsous CA, Shapiro L, Sheng Z, Huang Y, Ho DD: Increased resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7 to antibody neutralization. *bioRxiv* 2021. [Preprint] doi:[10.1101/2021.01.25.428137](https://doi.org/10.1101/2021.01.25.428137)
 - 80. Edara VV, Hudson WH, Xie X, Ahmed R, Suthar MS: Neutralizing antibodies against SARS-CoV-2 variants after infection and vaccination. *JAMA* 2021, 325:1896–1898
 - 81. Wu K, Werner AP, Moliva JI, Koch M, Choi A, Stewart-Jones GBE, Bennett H, Boyoglu-Barnum S, Shi W, Graham BS, Carfi A, Corbett KS, Seder RA, Edwards DK: mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. *bioRxiv* 2021. [Preprint] doi:[10.1101/2021.01.25.427948](https://doi.org/10.1101/2021.01.25.427948)
 - 82. Altmann DM, Boyton RJ, Beale R: Immunity to SARS-CoV-2 variants of concern. *Science* 2021, 371:1103–1104
 - 83. Wise J: Covid-19: the E484K mutation and the risks it poses. *BMJ* 2021, 372:n359
 - 84. Dagpunar J: Interim estimates of increased transmissibility, growth rate, and reproduction number of the Covid-19 B.1.617.2 variant of concern in the United Kingdom. *medRxiv* 2021. [Preprint] doi:[10.1101/2021.06.03.21258293](https://doi.org/10.1101/2021.06.03.21258293)