Identification and trajectory of growth of concerning SARS-CoV-2 variants in Houston, Texas, January through April 2021 based on 11,568 genome sequences


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, 6565 Fannin Street, Houston, Texas, 77030

Departments of Pathology and Laboratory Medicine, and Microbiology and Immunology, Weill Cornell Medical College, 1300 York Avenue, New York, New York, 10065

Consortium for Advanced Science and Engineering, 22 University of Chicago, 5801 South Ellis Avenue, Chicago, Illinois, 60637

Computing, Environment and Life Sciences, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, Illinois, 60439

Department of Molecular Biosciences and Institute of Molecular Biosciences, The University of Texas at Austin, Austin, Texas 78712

CCDC Army Research Laboratory-South, University of Texas, Austin, Texas 78712

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Running head (40 characters or less): SARS-CoV-2 concerning variants in Houston

¹Contributed equally.

# Address correspondence to James M. Musser, M.D., Ph.D., Department of Pathology and Genomic Medicine, Houston Methodist Research Institute, 6565 Fannin Street, Suite B490, Houston, Texas 77030. Tel: 713.441.5890, E-mail: jmmusser@houstonmethodist.org.

Disclosures: None.
Abstract

Genetic variants of the SARS-CoV-2 virus are of substantial concern because they can detrimentally alter the trajectory of the ongoing pandemic, and disease course in individual patients. Here we report genome sequences from 11,568 COVID-19 patients in the Houston Methodist healthcare system dispersed throughout the metroplex that were diagnosed from January 1, 2021 through April 30, 2021. This sample represents 94% of Houston Methodist cases and 4.6% of all reported cases in the metropolitan area during this period. The SARS-CoV-2 variant designated UK B.1.1.7 increased very rapidly, and now causes 75%-90% of all new cases in the Houston area. Five of the 2,543 B.1.1.7 genomes had an E484K change in spike protein. Compared with non-B.1.1.7 patients, individuals infected with B.1.1.7 had a significantly lower cycle threshold value (considered to be a proxy for higher virus load) and higher rate of hospitalization. Other variants (e.g., B.1.429, B.1.427, P.1, P.2, and R.1) also increased rapidly in frequency, although the magnitude was less than for B.1.1.7. We also identified 42 patients with a recently described R.1 variant that has an E484K amino acid replacement, and seven patients with the B.1.617 “India” variants. Vaccine breakthrough cases (n = 151) were caused by a heterogenous array of virus genotypes, including many that are not variants of interest or concern. In the aggregate, our study shows the occurrence of a diverse array of concerning SARS-CoV-2 variants circulating in a major metropolitan area, documents B.1.1.7 as the major cause of new cases in Houston and heralds the arrival and spread of B.1.617 variants in the metroplex.
[Introduction]

The global pandemic caused by SARS-CoV-2 that began in early 2020 has proved to be challenging for every academic health center and health system, hospital, and public health system in the United States and countries worldwide.\textsuperscript{1-7} The pandemic has also provided unprecedented opportunities for basic and translational research in all biomedical fields. We have systematically analyzed the molecular population genomics of SARS-CoV-2 in the ethnically and socioeconomically diverse metropolitan Houston area (population 7 million) since the first COVID-19 cases were reported in very early March 2020.\textsuperscript{8-11} Our studies are facilitated by a central molecular diagnostic laboratory that comprehensively identifies and retains all COVID-19 diagnostic specimens from our large healthcare system that includes eight hospitals, emergency care clinics, and outpatient centers distributed throughout the metropolitan region. In addition, we have leveraged our longstanding interest in pathogen genomics and sequencing infrastructure to investigate the spread of SARS-CoV-2 in metropolitan Houston.\textsuperscript{8-16} Among other discoveries, we have reported that the 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.\textsuperscript{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 5,085 genomes causing the first disease wave and massive second disease wave in Houston showed that all strains in the second wave had a Asp614Gly amino acid replacement in the spike protein.\textsuperscript{9} Importantly, this study was the first analysis of the molecular architecture of SARS-CoV-2 in two infection waves in any major metropolitan region. The
Asp614Gly polymorphism increases human transmission and infectivity \textit{in vitro} and \textit{in vivo} in animal infection models.\textsuperscript{17-22}

One key goal since the start of the pandemic has been to sequence all positive SARS-CoV-2 specimens and rapidly identifying 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 United States Centers for Disease Control and Prevention (CDC), World Health Organization (WHO), and Public Health England (PHE) (\url{https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html}, last accessed: May 16, 2021; \url{https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/}, last accessed: May 16, 2021; \url{https://www.gov.uk/government/collections/new-sars-cov-2-variant}, last accessed: May 16, 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 UK B.1.1.7 is of special interest because it has the ability to transmit very effectively, spread through populations rapidly, and has been reported to have a significantly higher mortality rate than non-B.1.1.7 infections (\url{https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563}, last accessed: May 17, 2020, \url{https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/947048/Technical_Briefing_VOC_SH_NJL2_SH2.pdf}, last accessed: May 17, 2020, \url{https://app.box.com/s/3lkcbxepqixkg4mv640dpvvg978ixjtf/file/756963730457}, last accessed: May 17, 2020, \url{https://cmmid.github.io/topics/covid19/uk-novel-variant.html}, last accessed: May

Based on sequencing 20,453 SARS-CoV-2 genomes causing COVID-19 disease in Houston, we recently reported that all VOIs and VOCs are circulating in the metropolitan region, making it the first community to document their presence.10 A follow-up study11 reported rapid increase of VOC UK B.1.1.7 in Houston; we estimated the variant had a doubling time of approximately 7 d. This rapid B.1.1.7 growth trajectory, raised the concern that this variant would cause nearly all new COVID-19 cases in metropolitan Houston by the end of March or early April 2021. This time frame is similar to an estimate made in late January by the CDC.34

Here we report data for 11,568 unique patients diagnosed between January 1, 2021 and April 30, 2021 infected with SARS-CoV-2, including 2,543 patients with the UK B.1.1.7 variant and 347 patients with infections caused by either B.1.429 or B.1.427, two closely related “California” VOC. We report that at the end of April, depending on the day, 75%-90% of all new cases of COVID-19 in metropolitan Houston were caused by B.1.1.7. Linked medical record information available for virtually all sequenced genomes permitted us to study the relationship between virus genotypes and patient phenotypes. Patients infected with B.1.1.7 had a significantly lower cycle threshold value in nasopharyngeal specimens (considered to be a proxy
for higher virus load) and 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. Five of the 2,543 B.1.1.7 genomes had an E484K change in spike protein that reduces binding by some neutralizing antibodies. Unexpectedly, we found five cases of B.1.1.7 in early December, resulting in a revised time frame for the introduction of this variant to Houston. We also identified seven patients with COVID-19 caused by B.1.617.1 or B.1.617.2, variants reported to be causing widespread disease and extensive public health concern in India.43-48 In the aggregate, our genome data show that VOC and VOI now account for the great majority of all new COVID-19 cases in our region.

Materials and Methods

Patient Specimens

Specimens were obtained from registered patients at Houston Methodist hospitals, associated facilities (e.g. urgent care centers), or 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 obtained from January 1, 2021 through April 30, 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 VOI and VOC. The study included 11,568 unique patients. 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 (EUA) from the FDA (https://www.fda.gov/medical-devices/emergency-situations-medical-devices/faqs-diagnostic-testing-sars-cov-2#offeringtests). 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, the Xpert Xpress SARS-CoV-2 test using Cepheid GeneXpert Infinity or Cepheid GeneXpert Xpress IV instruments, the cobas SARS-CoV-2 & Influenza A/B Assay using the Roche Liat system, the SARS-CoV-2 Assay using the Hologic Panther instrument, the Aptima SARS-CoV-2 Assay using the Hologic Panther Fusion system, the Cobas SARS-CoV-2 test using the Roche 6800 system, and the SARS-CoV-2 assay using Abbott Alinity m instruments. The great majority of tests were performed on material obtained from nasopharyngeal swabs immersed in universal transport media (UTM); oropharyngeal or nasal swabs, bronchoalveolar lavage fluid, or sputum treated with dithiothreitol (DTT) were sometimes used. Standardized specimen collection methods were used (https://vimeo.com/396996468/2228335d56).

SARS-CoV-2 Genome Sequencing

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

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

Viral genomes were assembled with the BV-BRC SARS-Cov2 assembly service (https://www.bv-brc.org/app/ComprehensiveSARS2Analysis). The One Codex SARS-CoV-2 variant calling and consensus assembly pipeline was used to assemble all sequences (https://github.com/onecodex/sars-cov-2.git) using default parameters and a minimum read depth of 3. Briefly, the pipeline uses seqtk version 1.3-r116 for sequence trimming (https://github.com/lh3/seqtk.git); minimap version 2.1 for aligning reads against reference genome Wuhan-Hu-1 (NC_045512.2); samtools version 1.11 for sequence and file manipulation; and iVar version 1.2.2 for primer trimming and variant calling. Genetic lineages, VOC, and VOI were identified based on genome sequence data and designated by Pangolin v. 2.4.2 with pangoLEARN module 2021-04-28 (https://cov-lineages.org/pangolin.html).

Patient Metadata and Geospatial Analysis

Patient metadata were acquired from the electronic medical record by standard informatics methods (Table 1). 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 (https://www.tableau.com/).

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 (https://covid-harriscounty.hub.arcgis.com/pages/cumulative-data). The third wave of COVID-19 started in approximately early November, following a prolonged disease trough occurring after the second wave (Figure 1). We studied 11,568 patients from January 1, 2021 through April 30, 2021, a period during which most of the variants were initially identified in Houston and several of them increased substantially (Figure 2, Table 1, Supplemental Figure 1).

The median age of the patients studied was 53.0 years and 53% were female; 5,534 (47.8%) of the 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 4.6%.

Occurrence of VOI and VOC

The CDC has identified eight VOI (B.1.525, B.1.526, B.1.526.1, P.2, B.1.617, B.1.617.1, B.1.617.2, B.1.617.2) and five VOC (B.1.1.7, P.1, B.1.351, B.1.427, and B.1.429) based on heightened concern about potential or proven threat to public health and individual patients. The following VOI were identified in our comprehensive sample of 11,568 genome sequences:

B.1.525 (n = 23), B.1.526 (n = 32), B.1.526.1 (n = 7), P.2 (n = 69), and B.1.617.1 (n = 5), B.1.617.2 (n = 2) All five VOC were found, including B.1.1.7 (n = 2,543), P.1 (n = 50), B.1.351 (n = 4), B.1.427 (n = 66), and B.1.429 (n = 281) (Figure 2, Supplemental Figure 1, Supplemental Figure 2; Table 1). B.1.1.7 rapidly increased and now dominates the new-
infection landscape in Houston (Figure 2). By the end of April, the B.1.1.7 variant caused 75%-90% of all new COVID-19 cases. In addition, we found that cases caused by variants P.1, P.2, and B.1.429 also increased during the study period, although not to the magnitude of B.1.1.7 infections (Figure 2).

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 country43-45 (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 VOI by the CDC. Variant B.1.617 is resistant to the monoclonal antibody Bamlanivimab (LY-Cov555), as assessed by an in vitro host-cell entry assay,46 and B.1.617.1 has been reported to be highly virulent in hamsters following intranasal inoculation.45 These two variants are characterized by a core group of four amino acid replacements in spike protein: L452R, E484Q, D614G, and P681R. We identified five patients infected with the B.1.617.1 variant, including two in March 2021, and three in April. We identified two patients with the B.1.617.2 variant in April 2021. Two of the patients with B.1.617.1 had a recent travel history to a high-prevalence country. Among these seven B.1.617.1 and B.1.617.2 variant samples, we also found five additional changes in spike protein, including del69-70, T95I, G142D, E154K, Q1071H, and H1101D. Based on the combination of spike amino acid changes, there were four distinct B.1.617.1 variants (del69-70, L452R, D614G, P681H; T95I, G142D, E154K, L452R, E484Q, D614G, P681R, Q1071H; G142D, E154K, L452R, E484Q, D614G, P681R Q1071H, H1101D; and G142D, E154K, L452R, E484Q, D614G, V615I, P681R Q1071H and H1101D), and two

The N440K amino acid change in spike protein has recently been of 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.49, 50 We identified five patients with this replacement, and all had the identical combination of spike amino acid replacements: L18R, T95I, R158S, N440K, D614G, P681H, A688V, S735A, T1027I. Pangolin categorized these strains as B.1. These five individuals were from three separate zip codes dispersed throughout metropolitan Houston (data not shown). Four of the five patients required hospitalization, and all were subsequently discharged.

Cycle Threshold (Ct) Value Comparison of B.1.1.7 and Non-B.1.1.7 Samples

Early in the pandemic it was reported9, 17 that nasopharyngeal samples from patients infected with strains having the spike protein 614Gly variant have, on average, significantly lower Ct values (considered to be a proxy for higher virus loads) on initial diagnosis. Most authorities think that higher virus load in the upper respiratory tract is related to ability to efficiently spread and infect others, although there are many factors that contribute to transmission and disease. We first tested the hypothesis that specimen from patients with B.1.1.7 infections had lower Ct values compared to non-B.1.1.7 patients based on data generated by the Abbott Alinity m or Hologic Panther molecular diagnostic assays. Consistent with the hypothesis, patient samples with the B.1.1.7 variant had significantly lower mean Ct value (Table 1 and Figure 3) on these
instruments. We next tested the hypothesis that other VOC and VOI have significantly lower Ct values. For this analysis, we removed B.1.1.7 samples because their inclusion would confound the data. The data show that B.1.429/B.1.427 samples also had significantly lower Ct values; further analysis found that this signal was attributable to the results for the B.1.429 samples (Figure 3). Ct data for the P.2 and R.1 patient samples were also significantly lower (Figure 3).

Taken together, these observations are consistent with the idea that, on average, several common SARS-CoV-2 variants have significantly lower Ct values, a feature that may make them better able to disseminate and become abundant. The sample sizes for the other VOI and VOIs are not adequate to analyze meaningfully.

Variant Geospatial Distribution

We next examined the geospatial distribution all VOI and VOC to investigate their extent of distribution 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 (due to small sample sizes), patients infected with all other variants were dispersed throughout metropolitan Houston, a finding consistent with the propensity of SARS-CoV-2 to spread rapidly between individuals (Figure 2 and Supplemental Figure 1).

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. We identified 284 samples with changes at E484 (E484K, n = 276; E484Q, n = 7; and E484D, n = 1) that occurred in many genetically diverse SARS-CoV-2 lineages, some of which have not shared a recent
common ancestor. For example, we found the E484K polymorphism in samples from 69 patients infected with VOI P.2 and 42 patients with newly described variant R.1.\textsuperscript{51-53} R.1 has the following core spike protein amino acid changes: W152L, E484K, D614G, and G769V (\url{https://outbreak.info/situation-reports?pango=R.1}, last accessed May 17, 2021). Some R.1 variants we identified also contain R21T, L54F, S254P, or P1162L changes. Of note, we identified five patients infected with B.1.1.7 plus the E484K amino acid change, and one patient each infected with B.1.1.7 sample containing either an E484Q or an E484D amino acid change. E484K replacement alters the immunologic profile of SARS-CoV-2,\textsuperscript{39, 40, 54-56} and Greaney et al.\textsuperscript{57} reported that E484Q reduced viral neutralization for some plasma samples.

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

In work conducted contemporaneously with the present study, we have routinely sequenced all genomes from earlier in the pandemic in Houston, including the uptick part of the third wave of disease occurring in November and December 2020 (\textbf{Figure 1}). We identified five patients in the first 10 days of December with infections caused by B.1.1.7, an unexpected result because the first Methodist patient previously documented with this VOC was identified in early January 2021,\textsuperscript{10} and the first Texas patient was announced by state public health authorities on January 7, 2021 (\url{https://www.dshs.texas.gov/news/releases/2021/20210107a.aspx}). Thus, our genome data revise this timeline. The first Houston, Texas B.1.1.7 patient was diagnosed in early December 2020, approximately one month earlier than previously known. Based on genome sequences deposited in GISAID (\url{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 US, a finding that further highlights the importance of comprehensive genome sequencing of large populations from metropolitan areas with diverse patient populations.

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.\textsuperscript{58-60} Two mRNA-based vaccines given Emergency Use Authorization by the Food and Drug Administration have been used widely in the United States (\url{https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine}, last accessed May 28, 2020, \url{https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/moderna-covid-19-vaccine}, last accessed May 28, 2020). Despite very high efficacy of the mRNA-based vaccines, a relatively small percentage of individuals who have received all recommended doses have developed either asymptomatic or symptomatic SARS-CoV-2 infections.\textsuperscript{61-63} The contributors to vaccine breakthroughs are not fully understood, but there is concern that genetic variants of SARS-CoV-2 may play a role, especially those with structural changes in spike protein that can alter immunologic characteristics. Consistent with this idea, McEwen et al.\textsuperscript{63} recently reported that all 20 vaccine breakthrough cases identified at the University of Washington were caused by VOC. To test the hypothesis that VOC and VOI were overrepresented among post-vaccination breakthrough infections among our COVID-19 cases, we examined metadata available for the 11,568 patients reported here. We identified 160 patients who met the criteria of vaccine breakthrough (i.e., infection occurring greater than 14
days after full vaccination was completed). SARS-CoV-2 genome sequence data were available for 151 cases. The patients were infected with a heterogenous array of variants only some of which were VOI or VOC (Figure 4). In the aggregate, there was a significant increase in VOI or VOC among the breakthrough cases ($p<0.001$). 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 4). 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 4). Of note, four cases in April were caused by VOC P.1; this variant was also increasing in disease frequency during this period (Figure 2).

**Discussion**

We analyzed the molecular population genomics of SARS-CoV-2 occurring in metropolitan Houston, Texas, with a focus on infections occurring early in 2021, from January 1 through April 30. Our study was based on genome sequences from 11,568 ethnically, socioeconomically, and geographically diverse patients distributed throughout the metropolitan area. We discovered that infections caused by UK B.1.1.7 increased very rapidly, and at the end of April caused 75-90% of all new cases in Houston. Compared with non-B.1.1.7 patients, individuals infected with B.1.1.7 had significantly lower virus Ct values and a higher rate of hospitalization, but no difference in length of stay or mortality. We also identified seven patients infected with B.1.617-family variants, genotypes now causing extensive disease in India.43-45
A key finding from our study was the very rapid trajectory of VOC B.1.1.7 in metropolitan Houston, an area with a population size of approximately 7 million. Several investigators have reported previously that patients infected with the B.1.1.7 VOC have significantly lower Ct values on initial diagnosis, but this has not been a universal finding.\textsuperscript{11, 64–69}

In the absence of quantitative virus cultures, the Ct value is viewed by many as a convenient proxy for virus load. We found (\textbf{Table 1, Figure 3}) a significantly decreased Ct value in nasopharyngeal swabs taken from B.1.1.7 patients compared to non-B.1.1.7 patients, a result consistent with prior reports.\textsuperscript{11, 65, 70–72} Our data are consistent with the potential for enhanced transmissibility of B.1.1.7. However, it is clear that there is no uniform relationship between Ct value and ability to disseminate. For example, we identified patients infected with B.1.1.7 who had high Ct values and non-B.1.1.7 patients with low Ct values. 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 capacity of virus genotypes to survive and be successfully transmitted. Collectively, our findings stress the need for more information about the relationship between Ct values, quantitative virus cultures, and specific genotypes of SARS-CoV-2.

We identified a significantly increased hospitalization rate for patients with B.1.1.7, compared to non-B.1.1.7 patients, but no significant difference in length of hospitalization or 28-day mortality (\textbf{Table 1}). Several studies\textsuperscript{23–37} have examined the relationship between disease severity and B.1.1.7. Patone et al.\textsuperscript{36} estimated the risk of critical care admission and overall mortality associated with B.1.1.7 compared to the original variant circulating in the UK among very large groups of patients. They reported that patients infected with B.1.1.7 have significantly increased risk for critical care admission and mortality compared to 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 caused more severe disease.

In the UK, at the end of April, the B.1.1.7 variant accounted for 98% of all COVID-19 cases (https://en.wikipedia.org/wiki/Lineage_B.1.1.7).27 A similar rapid increase in B.1.1.7 and population dominance has been reported in many countries, including Israel, France, Denmark, Norway, Lebanon, Norway, and other countries (https://en.wikipedia.org/wiki/Lineage_B.1.1.7). Our data show that this variant increased rapidly in metropolitan Houston and by the end of April caused 75%-90% of new COVID-19 cases. However, the increase in B.1.1.7 as percent 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 percent of cases caused by B.1.1.7 is attributed to the capacity of this variant to transmit more rapidly than other variants. We cannot rule out a contribution of a small but significant ability of B.1.1.7 to evade immunity induced by either natural infection or vaccination, and our data are consistent with this idea (Table 1). In this regard, data have been published showing that B.1.1.7 differs in some immunologic characteristics compared to “wild-type” SARS-CoV-2.73-80

SARS-CoV-2 variants with the E484K amino acid replacement are of particular concern in many areas including Brazil, South Africa, India and elsewhere (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, we identified the E484K change in several genetically distinct lineages of the virus, a finding likely
due to convergent evolution, as noted previously by others.\textsuperscript{39, 40, 54-56, 81} In the U.K., genome sequencing efforts have identified the E484K change in some B.1.1.7 samples, although it remains a minor subpopulation (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959426/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5.pdf, last accessed May 17, 2021).\textsuperscript{82} The B.1.1.7 plus E484K variant has been reported very infrequently elsewhere in the US (https://outbreak.info/situation-reports?pango=B.1.1.7&muts=S%3AE484K, last accessed May 17, 2021), and consistent with that, we found this amino acid change in only five of the 2,543 B.1.1.7 patients.

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).\textsuperscript{51-53} Cavanaugh et al.\textsuperscript{51} 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 variant R.1 was identified in mid-December 2020 and its prevalence increased during the study period (\textbf{Figure 2}).

We studied 151 patients who were documented to be fully immunized and had a specimen taken for diagnosis equal to or greater than 14 days after receiving their second dose of either the Pfizer ($n = 131$ patients, 87\%) or Moderna ($n = 20$ patients, 13\%) mRNA vaccine. Because we are sequencing the virus genomes causing the vast majority of Houston Methodist Hospital cases, we were able to discover that these vaccine breakthrough cases were caused by many diverse genotypes, including VOI, VOC, and many variants that are not assigned to these two important categories (Figure 4). Our results contrast with those reported recently by McEwen et al.\textsuperscript{63} in which all 20 breakthrough cases were caused by VOC. Although we found
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, genetically related members of the
B.1.617 variant family are thought to be contributing to the massive COVID-19 disease surge in
India. In this regard, the identification of patients in the Houston metropolitan area infected with
one of the known variants (B.1.617.1) and the closely related B.1.617.2 VOI is disconcerting.
One of the 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 isolates identified prior to this
starting on February 25th (www.gisaid.org, last accessed: May 17, 2021). After our study period
closed, we identified additional patients with variant B.1.617.1 or B.1.617.2, including some
with a very recent travel history outside the US. The B.1.617-family variants have amino acid
changes in spike protein 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.43-45, 47, 48 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 these and other related variants, and assess
if new variants that arise have biomedically relevant phenotypes.

Limitations

Our study has several limitations. During the January 1 through April 2021 study period,
253,756 cases of COVID-19 were reported in Harris County and eight contiguous counties
Thus, although we have sequenced 94% of all Houston Methodist cases identified during this period, our genome sample represents only 4.6% of all reported cases in the metropolitan region. Our eight hospitals and outpatient clinics are geographic 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 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 our study, although overall the number of cases in this age group is relatively small. Finally, virtually all SARS-CoV-2 genomes that we sequenced were obtained from symptomatic patients. Thus, our sample may underrepresent genotypes causing only asymptomatic carriage.

To summarize, by the end of April 2021, 75-90% of all new COVID-19 cases among ethnically, geographically, and socioeconomically diverse Houston Methodist health care patients were caused by the B.1.1.7 variant. Vaccine breakthrough cases are caused by SARS-CoV-2 that are genetically very diverse. Identification of the B.1.617 family of variants and B.1.1.7 plus E484K in metropolitan Houston is cause for concern. Inasmuch as our sample represents only 4.6% of all reported COVID-19 cases in the Houston area, by extrapolation it is reasonable to think that B.1.617-family variants have caused approximately 150 cases in our area. The rate and extent of spread of these variants should be monitored very closely by rapid genome sequence, coupled with linkage to patient metadata, such as disease severity. This is an
especially pressing issue for B.1.617-family variants because they are now beginning to become abundant and apparently outcompete B.1.1.7 in many areas of the UK


Acknowledgments

We thank the many talented and dedicated molecular technologists, and volunteers in the Molecular Diagnostics Laboratory and Methodist Research Institute for their dedicated efforts. We are indebted to Drs. Marc Boom and Dirk Sostman for their support, to generous Houston philanthropists for their support and to the Houston Methodist Academic Institute Infectious Diseases Fund that have made this ongoing project possible. James J. Davis and Robert Olson were funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. 75N93019C00076. We thank Jessica W. Podnar and personnel in the University of Texas Genome Sequencing and Analysis Facility for sequencing some of the genomes in this study. We thank Trina Trinh, Hung-Che Kuo and G. Nguyen for genome sequencing support. We gratefully acknowledge the originating and submitting laboratories of the SARS-CoV-2 genome sequences from GISAID’s EpiFlu™ Database used in some of the work presented here. We also thank 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., I.J.F, and J.G. performed research. All authors contributed to writing the manuscript.

Data availability: All genomes have been submitted to GISAID (www.gisaid.org)

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

Figure 2. Cumulative increase in SARS-CoV-2 variants and their distribution in metropolitan Houston. The time frame used is January 1, 2021 through April 30, 2021. The left column shows the cumulative increase in unique patients with each variant. The right column shows 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. Cycle threshold (Ct) for every SARS-CoV-2 patient sample tested using the Abbott Alinity m (panels A – F) or Hologic Panther (panel G – L) assays as described in Materials and Methods. Data are presented as median with 95% Confidence Interval. Mann-Whitney test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.
**Figure 4.** 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 151 vaccine breakthrough cases was identified in patient receiving either the Pfizer ($n = 131; \ 87\%$) or Moderna ($n = 20; \ 13\%$) vaccine.

**Supplemental Figure 1.** Four low-abundance SARS-CoV-2 variants and their geographic distribution in metropolitan Houston.

**Supplemental Figure 2.** Schematic showing structural changes present in spike protein of the major SARS-CoV-2 variants identified in the study, including VOI, VOC, and variant R.1. S1-NTD, S1 domain-aminoterminal domain; S1-RBD, S1 domain-receptor binding domain; S1, S1 domain; S2, S2 domain. The figure is a modified version of one presented in Long et al.^{10}

**References**


http://dx.doi.org/10.1016/S0140-6736(20)30154-9

http://dx.doi.org/10.1038/s41586-020-2008-3


http://dx.doi.org/10.1016/j.ajpath.2021.03.004

http://dx.doi.org/10.1101/2021.03.16.21253753

http://dx.doi.org/10.1128/mBio.00489-17


http://dx.doi.org/10.1016/j.cell.2021.03.052


[68] Hay JA, Kennedy-Shaffer L, Kanjilal S, Lennon NJ, Gabriel SB, Lipsitch M, Mina MJ: Estimating epidemiologic dynamics from cross-sectional viral load distributions. medRxiv 2021:2020.10.08.20204222. http://dx.doi.org/10.1101/2021.05.20.21257552; this version posted May 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.


http://dx.doi.org/10.1126/science.abg6105
http://dx.doi.org/10.1136/bmj.n597
http://dx.doi.org/10.1001/jama.2021.4388
http://dx.doi.org/10.1136/bmj.n359
Table 1. Summary of pertinent patient metadata for the 11,568 unique patients

<table>
<thead>
<tr>
<th></th>
<th>B.1.1.7 Variant</th>
<th>Other Variants</th>
<th>Total</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. (%) with data</strong></td>
<td>2543 (22.0%)</td>
<td>9025 (78.0%)</td>
<td>11568</td>
<td></td>
</tr>
<tr>
<td><strong>Patient Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Age (Years)</td>
<td>50.0</td>
<td>54.0</td>
<td></td>
<td><strong>P=0.0001</strong></td>
</tr>
<tr>
<td>Female</td>
<td>1351 (53.1%)</td>
<td>4778 (52.9%)</td>
<td>6129 (53.0%)</td>
<td><strong>P=0.8749</strong></td>
</tr>
<tr>
<td>Male</td>
<td>1192 (46.9%)</td>
<td>4247 (47.1%)</td>
<td>5439 (47.0%)</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>984 (39.0%)</td>
<td>3666 (40.8%)</td>
<td>4650 (40.4%)</td>
<td><strong>P&lt;0.0001</strong></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>741 (29.4%)</td>
<td>2656 (29.6%)</td>
<td>3397 (29.5%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>581 (23.0%)</td>
<td>1606 (17.9%)</td>
<td>2187 (19.0%)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>116 (4.6%)</td>
<td>533 (5.9%)</td>
<td>649 (5.6%)</td>
<td></td>
</tr>
<tr>
<td>Native American</td>
<td>12 (0.5%)</td>
<td>27 (0.3%)</td>
<td>39 (0.3%)</td>
<td></td>
</tr>
<tr>
<td>Hawaiian/Pacific Islander</td>
<td>3 (0.1%)</td>
<td>23 (0.3%)</td>
<td>26 (0.2%)</td>
<td></td>
</tr>
<tr>
<td>Unavailable</td>
<td>86 (3.4%)</td>
<td>475 (5.3%)</td>
<td>561 (4.9%)</td>
<td></td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median BMI</td>
<td>30.4</td>
<td>29.5</td>
<td>n=(10164)</td>
<td><strong>P=0.0001</strong></td>
</tr>
<tr>
<td>&lt;18.5</td>
<td>32 (1.3%)</td>
<td>141 (1.6%)</td>
<td>173 (1.5%)</td>
<td></td>
</tr>
<tr>
<td>18.5-25</td>
<td>402 (15.8%)</td>
<td>1570 (17.4%)</td>
<td>1972 (17.0%)</td>
<td></td>
</tr>
<tr>
<td>25-30</td>
<td>655 (25.8%)</td>
<td>2455 (27.2%)</td>
<td>3110 (26.9%)</td>
<td></td>
</tr>
<tr>
<td>30-35</td>
<td>567 (22.3%)</td>
<td>1810 (20.1%)</td>
<td>2377 (20.5%)</td>
<td></td>
</tr>
<tr>
<td>&gt;=35</td>
<td>632 (24.9%)</td>
<td>1900 (21.1%)</td>
<td>2532 (21.9%)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>255 (10.0%)</td>
<td>1149 (12.7%)</td>
<td>1404 (12.1%)</td>
<td></td>
</tr>
<tr>
<td><strong>Admission Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admitted</td>
<td>1350 (53.1%)</td>
<td>4184 (46.4%)</td>
<td>5534 (47.8%)</td>
<td><strong>P=0.0001</strong></td>
</tr>
<tr>
<td>Not Admitted</td>
<td>1193 (46.9%)</td>
<td>4841 (53.6%)</td>
<td>6034 (52.2%)</td>
<td></td>
</tr>
<tr>
<td>Median LOS (Days)</td>
<td>5.1</td>
<td>5.2</td>
<td>n=(10164)</td>
<td><strong>P=0.9964</strong></td>
</tr>
<tr>
<td><strong>Max Respiratory Support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECMO</td>
<td>8 (0.6%)</td>
<td>17 (0.4%)</td>
<td>25 (0.5%)</td>
<td><strong>P=0.0074</strong></td>
</tr>
<tr>
<td>Mechanical Ventilation</td>
<td>116 (8.6%)</td>
<td>362 (8.7%)</td>
<td>478 (8.6%)</td>
<td></td>
</tr>
<tr>
<td>Non-Invasive Ventilation</td>
<td>120 (8.9%)</td>
<td>434 (10.4%)</td>
<td>554 (10.0%)</td>
<td></td>
</tr>
<tr>
<td>High Flow Oxygen</td>
<td>277 (20.5%)</td>
<td>680 (16.3%)</td>
<td>957 (17.3%)</td>
<td></td>
</tr>
<tr>
<td>Low Flow Oxygen</td>
<td>568 (42.1%)</td>
<td>1802 (43.1%)</td>
<td>2370 (42.8%)</td>
<td></td>
</tr>
<tr>
<td>Room Air</td>
<td>261 (19.3%)</td>
<td>889 (21.2%)</td>
<td>1150 (20.8%)</td>
<td></td>
</tr>
<tr>
<td><strong>Mortality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>2430 (95.6%)</td>
<td>8607 (95.4%)</td>
<td>11037 (95.4%)</td>
<td><strong>P=0.7475</strong></td>
</tr>
<tr>
<td>Deceased</td>
<td>113 (4.4%)</td>
<td>418 (4.6%)</td>
<td>531 (4.6%)</td>
<td></td>
</tr>
<tr>
<td><strong>Median PCR Cycle Threshold</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbott Alinity</td>
<td>23.4</td>
<td>26.8</td>
<td>n=(4171)</td>
<td><strong>P=0.0001</strong></td>
</tr>
<tr>
<td>Hologic Panther</td>
<td>24.5</td>
<td>26.2</td>
<td>n=(1838)</td>
<td><strong>P=0.0109</strong></td>
</tr>
<tr>
<td><strong>Vaccine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No vaccine</td>
<td>2358 (92.7%)</td>
<td>8556 (94.8%)</td>
<td>10914 (94.3%)</td>
<td><strong>P=0.0001</strong></td>
</tr>
<tr>
<td>&gt;7 days past 1st Vaccine</td>
<td>107 (4.2%)</td>
<td>396 (4.4%)</td>
<td>503 (4.3%)</td>
<td></td>
</tr>
<tr>
<td>&gt;14 days past 2nd Vaccine</td>
<td>78 (3.1%)</td>
<td>73 (0.8%)</td>
<td>151 (1.3%)</td>
<td></td>
</tr>
</tbody>
</table>

Data include 11568 unique patients with high-quality sequence results between January 1, 2021, and April 30, 2021. CI: confidence interval; ECMO: extracorporeal membrane oxygenation; LOS: length of stay.
Figure 2

B.1.1.7
\((n = 2543)\)

B.1.429
\((n = 281)\)

P.2
\((n = 69)\)

P.1
\((n = 50)\)
Figure 3
Figure 4

The figure shows the distribution of patients with variants of interest/concern over the months of January to April in 2021. The variants are color-coded as follows:

- **B.1**
- **B.1.1**
- **B.1.1.117**
- **B.1.1.222**
- **B.1.1.519**
- **B.1.177**
- **B.1.2**
- **B.1.234**
- **B.1.595**
- **B.1.596**
- **B.1.609**

The data is divided into two categories:

1. **Not Variant of Interest/Concern**
2. **Yes Variant of Interest/Concern**

Each category is further divided by month:
- **January**
- **February**
- **March**
- **April**

The total number of patients in each category is indicated:
- **(n=64)**
- **(n=87)**