Open access • Posted Content • DOI:10.1101/2021.02.26.21252227

Sequence Analysis of 20,453 SARS-CoV-2 Genomes from the Houston Metropolitan Area Identifies the Emergence and Widespread Distribution of Multiple Isolates of All Major Variants of Concern — Source link

Scott Wesley Long, Randall J. Olsen, Randall J. Olsen, Paul A. Christensen ...+16 more authors

Institutions: Houston Methodist Hospital, Cornell University, University of Chicago, Argonne National Laboratory ...+2 more institutions

Published on: 02 Mar 2021 - medRxiv (Cold Spring Harbor Laboratory Press)

Topics: Population

Sequence Analysis of 20,453 SARS-CoV-2 Genomes from the Houston Metropolitan Area Identifies the Emergence and Widespread Distribution of Multiple Isolates of All Major Variants of Concern

S. Wesley Long,[a,b,1] Randall J. Olsen,[a,b,1] Paul A. Christensen,[a] Sishir Subedi,[a] Robert Olson,[c,d] James J. Davis,[c,d] Matthew Ojeda Saavedra,[a] Prasanti Yerramilli,[a] Layne Pruitt,[a] Kristina Reppond,[a] Madison N. Shyer,[a] Jessica Cambric,[a] Ilya J. Finkelstein,[e] Jimmy Gollihar,[a,f] and James M. Musser [a,b#]

From the [a]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

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

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

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

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

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

22

23 Number of text pages: 12

24 Number of tables: 1

25 Number of figures: 2

26 Running head (40 characters or less): SARS-CoV-2 variants of concern in Houston, TX

27

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

31

32 Disclosures: None.

33

34

35

36

37

38

39

40

41

42

43

[Abstract (220 words)]

Since the beginning of the SARS-CoV-2 pandemic, there has been international concern about the emergence of virus variants with mutations that increase transmissibility, enhance escape from the human immune response, or otherwise alter biologically important phenotypes. In late 2020, several "variants of concern" emerged globally, including the UK variant (B.1.1.7), South Africa variant (B.1.351), Brazil variants (P.1 and P.2), and two related California "variants of interest" (B.1.429 and B.1.427). These variants are believed to have enhanced transmissibility capacity. For the South Africa and Brazil variants, there is evidence that mutations in spike protein permit it to escape from some vaccines and therapeutic monoclonal antibodies. Based on our extensive genome sequencing program involving 20,453 virus specimens from COVID-19 patients dating from March 2020, we report identification of all important SARS-CoV-2 variants among Houston Methodist Hospital patients residing in the greater metropolitan area. Although these variants are currently at relatively low frequency in the population, they are geographically widespread. Houston is the first city in the United States to have all variants documented by genome sequencing. As vaccine deployment accelerates worldwide, increased genomic surveillance of SARS-CoV-2 is essential to understanding the presence and frequency of consequential variants and their patterns and trajectory of dissemination. This information is critical for medical and public health efforts to effectively address and mitigate this global crisis.

[Introduction]

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of coronavirus disease 2019 (COVID-19). Since first being identified in December 2019,[1-4] the virus has spread globally and is responsible for massive human morbidity and mortality worldwide.[5-9] At the onset of the pandemic, effective treatments for COVID-19 were lacking. But as a result of intense global research efforts, monoclonal antibody (mAbs) therapies[10, 11] and several vaccines,[12, 13] primarily directed against the spike protein, have been developed to treat and prevent SARS-CoV-2 infection.

In late 2020 the international research community described several SARS-CoV-2 "variants of concern" that warranted special scrutiny. These include the United Kingdom (UK) variant (B.1.1.7), South Africa variant (B.1.351), Brazil variants (P.1 and P.2) and two California variants (B.1.429/CAL.20C and B.1.427/CAL.20C).[14-22] These virus variants were designated as "concerning" predominantly due to their reported enhanced person-to-person transmission in some geographic areas, and they have since been detected in several countries worldwide. For example, the UK B.1.1.7 variant spread rapidly in southeast England where it caused large numbers of COVID-19 cases,[14] and was identified shortly thereafter in the United States (US) [CDC; https://www.cdc.gov/coronavirus/2019-ncov/transmission/variant.html]. [23] More than 1,600 cases have since been documented in the US, and at least one large outbreak recently was reported in a Michigan prison.[24, 25] There is concern at the Centers for Disease Control and Prevention (CDC) that it could become the dominant variant causing disease in the US by March.[23, 24, 26] Moreover, the UK B.1.1.7 variant may be linked to an increased death rate compared to other virus types, adding further concern.[18, 21, 27, 28]

Similarly, the South Africa and Brazil variants caused large disease outbreaks in their respective countries.[19, 20] These variants also are of concern because they contain a mutation (E484K) in the spike protein that decreases efficacy of some therapeutic mAbs, decreases *in vitro* virus neutralization, and may result in potential escape from immunity induced by natural infection or vaccination.[29-37] All three variants (UK B.1.1.7, Brazil P.1, and South Africa B.1.351) also have a N501Y mutation in spike protein that is associated with stronger binding to the ACE2 receptor, possibly contributing to increased transmissibility.[38,39]

The Houston metropolitan area is the fourth largest and most ethnically diverse city in the US, with a population of approximately 7 million (https://www.houston.org/houston-data).[40] The 2,400-bed Houston Methodist health system has eight hospitals and cares for a large, multiethnic, and geographically and socioeconomically diverse patient population throughout greater Houston. The eight Houston Methodist hospitals have a single central molecular diagnostic laboratory, which means that all RT-PCR-specimens can readily be identified, banked, and subjected to further study as needed. In addition, the Department of Pathology and Genomic Medicine has a long-standing record of integrating genome sequencing efforts into clinical care and research, especially related to microbial pathogens infecting our patients.[41-49] In the aggregate, strategic co-localization of these diagnostic attributes coupled with a contiguous research institute building seamlessly facilitates comprehensive population genomic studies of SARS-CoV-2 viruses causing infections in the Houston metropolitan region.[46, 49]

Before the SARS-CoV-2 virus arrived in Houston, we planned an integrated strategy to confront and mitigate this microbial threat to our patients. In addition to rapidly validating an

RT-PCR test for the virus, we instituted a plan to sequence the genome of every positive specimen from patients within the Houston Methodist system, with the goal of understanding pathogen spread in our community and identifying biologically-important mutant viruses. We previously described the detailed population genomics of the first and second waves of SARS-CoV-2 in the Houston metropolitan region.[46,49] We have continued to sequence positive SARS-CoV-2 specimens with the goal of monitoring for variants of concern and genome mutations that may be associated with patient outcome or therapeutic failure.

This report describes the identification of multiple isolates of important SARS-CoV-2 variants, including the UK B.1.1.7, South Africa B.1.351, Brazil P.1 and P.2, and California B.1.429 and B.1.427 variants in Houston patient specimens collected from December 2020 through mid-February 2021. These findings represent the first detection of the South Africa and Brazil variants in Texas and only the second time UK variants have been identified in Houston. Greater Houston is the first metroplex in the US documented to have all of these important and concerning variants circulating among its residents. Our discoveries further illustrate the need for increased population genomic and epidemiology efforts to identify and help track dissemination of these variants, monitor development of new variants, and assess the relationship between variants and COVID-19 disease outcomes.

Materials and Methods

Patient Specimens

All specimens were obtained from individuals who were registered patients at Houston Methodist hospitals, associated facilities (e.g. urgent care centers), or institutions in the greater Houston metropolitan region that use our laboratory services. Virtually all individuals had signs or symptoms consistent with COVID-19 disease. This 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 high degree of 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). 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 SARS-CoV-2 Assay using the Hologic Panther instrument, the Aptima SARS-CoV-2 Assay using the Hologic Panther Fusion system and the SARS-CoV-2 assay using Abbott Alinity m instruments. All assays were performed according to the manufacturer's instructions. Testing was performed on material obtained from nasopharyngeal, oropharyngeal, or nasal swabs immersed in universal transport media (UTM), bronchoalveolar lavage fluid, or sputum treated with dithiothreitol (DTT). To standardize specimen collection, an instructional video was created for Houston Methodist healthcare workers (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). Long reads were generated with the LSK-109 sequencing kit, 24 native barcodes (NBD104 and NBD114 kits), and a GridION instrument (Oxford Nanopore). Short sequence reads were generated with either a NextSeq 550 or NovaSeq 6000 instrument (Illumina).

SARS-CoV-2 Genome Sequence Analysis

Viral genomes were assembled with the BV-BRC SARS-Cov2 assembly service (https://www.bv-brc.org/app/ComprehensiveSARS2Analysis).[50] The One Codex SARS-CoV-2 variant calling and consensus assembly pipeline was chosen for assembling 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[51] for aligning reads against reference genome Wuhan-Hu-1 (NC_045512.2); samtools version 1.11 for sequence and file manipulation[52]; and iVar version 1.2.2 for primer trimming and variant calling.[53]

Geospatial Analysis

The patient home address zip codes were used to visualize the geospatial distribution of spread for each variant of concern. Figures were generated using Tableau version 2020.3.4.

Results

Since the start of the SARS-CoV-2 pandemic, we have sequenced 20,453 specimens collected from patients in the Houston metropolitan area. In genome sequencing conducted in January and February 2021, we discovered our first variants of concern. These included 23 UK variants (B.1.1.7), two South African variants (B.1.351), and four Brazilian variants (P.1). We also identified 162 patients infected with the California variants (B.1.429, *N* = 143; B.1.427, *N* = 19) and 39 patients infected with Brazil P.2 variants 2020 (Table 1) .

UK Variant of Concern (B.1.1.7)

The UK variant known as B.1.1.7 was first identified in September 2020 in the UK and was designated as a variant of concern in South London on December 14, 2020. It was strongly associated with a resurgence of SARS-CoV-2 infections in that region and rapidly became the dominant lineage.[26] Importantly, the UK has the most extensive SARS-CoV-2 genome sequencing program in the world, making them particularly well situated to rapidly identify new variants. Of the ~500,000 SARS-CoV-2 genome sequences submitted to GISAID from global sources, approximately one-half originated from collaborating laboratories in the UK as part of the COVID-19 Genomics UK Consortium.[54, 55]

The UK B.1.1.7 variant is of particular concern because it has an unusually large number of genome mutations, including multiple changes in spike protein (Figure 1). Some of the mutations of primary concern include N501Y located in the receptor binding domain, and a two amino acid deletion (del69-70) that has arisen in multiple SARS-CoV-2 genetic backgrounds and

is associated with increased transmissibility[26]. In addition, evidence has been presented from the UK that B.1.1.7 strains may cause increased hospitalization and mortality.[18, 21, 27, 56] The first patient we identified in Houston with a B.1.1.7 variant was diagnosed the second week of January, 2020; thus far we have identified 23 patients with this variant of concern (Table 1). Of note, none of our first three patients had an international travel history, suggesting that they acquired the B.1.1.7 infections either locally or during domestic travel. Preliminary evidence indicates that immune sera from the Pfizer-BioNTech SARS-CoV-2 vaccine retain the ability to neutralize B.1.1.7 variants *in vitro*.[57] Additional studies have found that convalescent plasma from many patients, and some monoclonal antibody therapies, retain the ability to neutralize B.1.1.7 variant SARS-CoV-2 *in vitro*.[34, 35]

South Africa Variant of Concern (B.1.351)

The South Africa B.1.351 variant of concern was first identified in a COVID-19 epidemic wave occurring in Nelson Mandela Bay in October 2020.[19] This variant was concerning because of its large number of spike protein mutations (including K417N, E484K, and N501Y) (Figure 1) and apparent increased transmissibility.[19, 38] These three mutations are located in the receptor binding domain of spike and may decrease the effectiveness of some mAb therapies and vaccines.[29-31, 34, 35, 58] The first South Africa variant detected in Houston was identified in a patient specimen we collected the last week of December, 2020, and the second patient's specimen was collected the first week of January, 2021. Of note, these Houston Methodist Hospital patients had no known international travel history, suggesting domestic acquisition of this B.1.351 variant.

Brazil Variants of Concern (P.1 and P.2)

The P.1 variant of concern was reported to have originated in Manaus, Brazil, and like the South Africa B.1.351 variant, has numerous mutations in spike protein, including E484K and N501Y (Figure 1).[59] We identified our first P.1 variant in Houston specimens the third week of January, 2021. In total, we have identified four P.1 variants in our patient samples (Table 1). The P.2 variant began to spread in Brazil in earnest in October of 2020, similar to P.1.[60, 61] It also has a E484K amino acid change in the RBD of spike protein (Figure 1), similar to variant P.1 and B.1.351.[17] We first identified a P.2 variant in a patient specimen obtained the last week of December, 2020. In total, we have documented 39 P.2 variants in our patient specimens (Table 1).

California Variants (B.1.429 and B.1.427)

The emergence of what became known as the California variant, originally known as CAL.20C and later designated as lineages B.1.429 and B.1.427, was first identified in Los Angeles County in July 2020 as a single isolate.[62, 63] This variant re-emerged in October 2020 and was associated with an increasing number of cases during a wave of SARS-CoV-2 infections in the region.[16] Variant B.1.429 accounted for 36% of isolates collected from late November to late December 2020 in Los Angeles County.[16] Since November 2020, this variant has been detected in 42 states in the US,[63] and was first found in Houston Methodist Hospital patients in specimens obtained the last week of December, 2020. We identified 143 and 19 patients with the B.1.429 and B.1.427 isolates, respectively (Table 1). The B.1.427 variant is closely related to B.1.429 (Figure

1) and has spread from California to 34 states since October 2020.[62] The California variants are noteworthy primarily for their emergence and very rapid spread in Los Angeles County and identification elsewhere in the US. However, as of February 17, 2021, they have not been designated as variants of concern by the Centers for Disease Control.

Geospatial Distribution of Variants

Given the importance of the identification of these SARS-CoV-2 variants in the Houston metropolitan area, we examined their geospatial distribution to investigate the extent of dissemination (Figure 2). With the exception of the B.1.351 variant, patients infected with all other variants resided in widely dispersed areas of the metropolitan area. This finding is consistent with the well-known propensity of SARS-CoV-2 to spread rapidly between individuals, and especially so for these variants of concern [19, 23, 24, 27, 64-66].

Discussion

Here we report discovery of the UK (B.1.1.7), South Africa (B.1.351), and Brazil (P.1) SARS-CoV-2 variants of concern from patients in the Houston metropolitan region. We also identified geographically-widespread dissemination of the Cal.20C California (B.1.429 and B.1.427) variants of interest. These four SARS-CoV-2 variants are distributed across a large geospatial region in the metropolitan region (Figure 2), indicating successful patient-to-patient transmission among Houstonians. None of the affected patients were from a common household or reported recent international travel, suggesting that every infection was independently acquired locally or during domestic travel. Given that Houston is a culturally- and

ethnically-diverse population center with two international airports, a major shipping center, and a global energy sector, the discovery of patients infected with each of the four concerning SARS-CoV-2 variants is not unexpected but it is disquieting. With this report, Houston now becomes the first US city to document patients infected with each of the four SARS-CoV-2 variants of concern or interest, testament to our aggressive sequencing of COVID-19 patient samples.

The P.2 variant gained recent attention in the scientific and lay press because it has been reported to cause SARS-CoV-2 reinfections.[67, 68] We identified 39 P.2 infections among Houston patients. Although it is currently a numerically minor cause of all Houston-area infections, P.2 is now the most common SARS-CoV-2 variant of concern in our population.

The E484K amino acid replacement in spike protein is characteristic of P.1, P.2, and B.1.351 strains (Figure 1). It has independently arisen in many different SARS-CoV-2 genomic backgrounds, including some B.1.1.7 strains.[69] This amino acid replacement has caused substantial public health concern due to its potentially detrimental effects on neutralizing activity of therapeutic mAbs, sera obtained from naturally infected individuals, and post-vaccination sera.[70, 71] That is, the E484K amino acid change may facilitate vaccine escape. Among our Houston SARS-CoV-2 genomes, E484K was detected 84 times (0.4% of the total genomes sequenced). It was first detected in a respiratory specimen collected in July 2020, near the peak of our second massive wave of infections,[46] and has been identified in many diverse genomic backgrounds thereafter. Due to this strong signal of convergent evolution, we will continue to closely monitor all Houston SARS-CoV-2 genomes for the E484K amino acid change.

283 Recently, the Q677H amino acid change in spike protein has been identified in SARS-
284 CoV-2 patient samples collected in multiple US states and other global locations.[72, 73] Q677H has
285 arisen in at least six distinct genomic backgrounds.[73] A Q667P amino acid change has also been
286 identified.[73] Among the Houston genomes, Q677H occurred 288 times (1.4%) and is encoded
287 by two different nucleotide changes. We also identifed two other amino acid changes, 677P (in
288 330 genomes, 1.6%) and Q677K (2 genomes, <0.1%) in Houston. Taken together, these data
289 suggest selection for a yet to be determined biologic phenotype associated with amino acid
290 replacements at position 677.
291 Many population genomic studies performed in varous global locations have clearly
292 demonstrated that SARS-CoV-2 variants with biologically-relevant phenotypes have evolved.
293 Emergence of new variants underscores the need for ongoing extensive genomic sequencing
294 efforts for early identification and public health warning. In support of these efforts, our
295 laboratory has devoted substantial resources to SARS-CoV-2 genomics, resulting in sequence
296 analysis of more genomes than any other state in the US.[54] Since March 2020, approximately
297 36,500 SARS-CoV-2 positive patients have received care in our Houston Methodist health
298 system, and we have sequenced 20,453 virus genomes. In total, this dataset represents 56% of
299 our Houston Methodist COVID-19 patients. Inasmuch as almost 500,000 COVID-19 infections
300 have been reported in the Houston metropolitan area,[74] we have sequenced the genome of
301 4.1% of all cases reported in our area. Based on modeling, this sample depth may be sufficient
302 to identify all variants occurring at a biologically-relevant frequency.[75] Due to the very wide
303 geographic catchment of our eight-hospital system that serves a very diverse patient
304 population, the data presented here likely reflect a reasonably detailed overview of SARS-CoV-2

genomic diversity throughout our metroplex. This comparatively deep sampling of the Houston metropolitan SARS-CoV-2 population enabled us to identify patients infected with variants of concern, and provided information regarding the timeframe of initial presence and frequency of each variant. We modeled our strategy on the aggressive genome sequencing being conducted in the UK, a global leader in SARS-CoV-2 genome sequencing.[76]

Our large SARS-CoV-2 genome dataset and comprehensive infrastructure are unique resources. By linking the SARS-CoV-2 whole genome sequence data to patient metadata present in our electronic medical record, we are able to use analytic tools such as high-performance compute clusters and machine learning to investigate the relationship between genomic diversity and phenotypic traits such as strain virulence or patient outcomes.[46] For example, recent reports of increased mortality caused by B.1.1.7 variant strains are very concerning and worthy of further investigation.[18, 21, 27, 28] Similarly, our COVID-19 biobank has cryopreserved respiratory samples, white blood cells, serum, plasma, and formalin-fixed paraffin-embedded tissues for use in downstream investigations such as viral neutralization assays, RNA sequencing, and immune repertoire analysis.

Our goal is to sequence the SARS-CoV-2 genome of every infected patient in our health care system in near-real time, and expand outward to other patients in our community. Consistent with these goals, the American Rescue Plan announced by the Biden administration proposes to substantially fund sequencing capacity in the US. However, it remains unclear how these important funds will be distributed.[77] Our results from a major metropolitan region in the US underscore the necessity of greatly increased genome surveillance to rapidly identify and track the emergence and introduction of SARS-CoV-2 variants in the US and local areas.

Acknowledgments

We thank Drs. Jessica Thomas and Zejuan Li, Erika Walker, the many very talented and dedicated molecular technologists, and many labor pool 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 many very generous Houston philanthropists for their tremendous support and to the Houston Methodist Academic Institute Infectious Diseases Fund that have made this ongoing project possible. James 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 acknowledge the assistance of Maulik Shukla and Marcus Nguyen at PATRIC Bioinformatics Resource Center for their assistance with sequence analysis. 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

348 improvements, and Dr. Kathryn Stockbauer, Sasha Pejerrey, Adrienne Winston, and Heather
349 McConnell for help with figures, tables, and editorial contributions.

350

351 Author Contributions

352 J.M.M. conceptualized and designed the project; S.W.L, R.J.O., P.A.C., S.S., R.O., J.J.D., M.S., P.Y.,
353 L.P., K.R., M.N.S, J.C., I.J.F, and J.G. performed research. All authors contributed to writing the
354 manuscript.

355

356

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

358

359

360

References

[1] 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.

[2] 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: A Novel Coronavirus from Patients with Pneumonia in China, 2019. New England Journal of Medicine 2020, 382:727-33.

[3] Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi 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, 395:514-23.

[4] 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, 579:265-9.

[5] World Health Organization Coronavirus Disease 2019 (COVID-19) Situation Report. 2020.

[6] Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, Penzar D, Perlman S, Poon LLM, Samborskiy DV, Sidorov IA, Sola I, Ziebuhr J, Coronaviridae Study Group of the International Committee on Taxonomy of V: The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology 2020, 5:536-44.

[7] Wang C, Horby PW, Hayden FG, Gao GF: A novel coronavirus outbreak of global health concern. Lancet 2020, 395:470-3.

[8] Perlman S: Another Decade, Another Coronavirus. New England Journal of Medicine 2020, 382:760-2.

[9] Allel K, Tapia-Muñoz 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 2020:1-14.

[10] Chen P, Nirula A, Heller B, Gottlieb RL, Boscia J, Morris J, Huhn G, Cardona J, Mocherla B, Stosor V, Shawa I, Adams AC, Van Naarden J, Custer KL, Shen L, Durante M, Oakley G, Schade AE, Sabo J, Patel DR, Klekotka P, Skovronsky DM, Investigators B-: SARS-CoV-2 Neutralizing Antibody LY-CoV555 in Outpatients with Covid-19. N Engl J Med 2021, 384:229-37.

[11] Weinreich DM, Sivapalasingam S, Norton T, Ali S, Gao H, Bhore R, Musser BJ, Soo Y, Rofail D, Im J, Perry C, Pan C, Hosain R, Mahmood A, Davis JD, Turner KC, Hooper AT, Hamilton JD, Baum A, Kyratsous CA, Kim Y, Cook A, Kampman W, Kohli A, Sachdeva Y, Graber X, Kowal B, DiCioccio T, Stahl N, Lipsich L, Braunstein N, Herman G, Yancopoulos GD, Trial I: REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19. N Engl J Med 2021, 384:238-51.

[12] Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, Diemert D, Spector SA, Rouphael N, Creech CB, McGettigan J, Khetan S, Segall N, Solis J, Brosz A, Fierro C, Schwartz H, Neuzil K, Corey L, Gilbert P, Janes H, Follmann D, Marovich M, Mascola J, Polakowski L, Ledgerwood J, Graham BS, Bennett H, Pajon R, Knightly C, Leav B, Deng W, Zhou H, Han S, Ivarsson M, Miller J, Zaks T, Group CS: Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. N Engl J Med 2021, 384:403-16.

403 [13] Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Perez Marc G,
404 Moreira ED, Zerbini C, Bailey R, Swanson KA, Roychoudhury S, Koury K, Li P, Kalina WV, Cooper
405 D, Frenck RW, Jr., Hammitt LL, Tureci O, Nell H, Schaefer A, Unal S, Tresnan DB, Mather S,
406 Dormitzer PR, Sahin U, Jansen KU, Gruber WC, Group CCT: Safety and Efficacy of the BNT162b2
407 mRNA Covid-19 Vaccine. N Engl J Med 2020, 383:2603-15.
408 [14] Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, Connor T, Peacock T,
409 Robertson DL, Volzon E, (CoG-UK) C-GCU: Preliminary genomic characterisation of an emergent
410 SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. 2020.
411 [15] Naveca F, Nascimento V, Souza V, Corado A, Nascimento F, Silva G, Costa Á, Duarte D,
412 Pessoa K, Gonçalves L, Brandão MJ, Jesus M, Fernandes C, Pinto R, Silva M, Mattos T, Wallau
413 GL, Siqueira MM, Resende PC, Delatorre E, Gräf T, Bello G: Phylogenetic relationship of SARS-
414 CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K
415 and N501Y in the Spike protein. 2021.
416 [16] Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E: Emergence of a Novel
417 SARS-CoV-2 Variant in Southern California. JAMA 2021.
418 [17] Voloch CM, Silva F Rd, de Almeida LGP, Cardoso CC, Brustolini OJ, Gerber AL, Guimarães
419 APdC, Mariani D, Costa RMd, Ferreira OC, Cavalcanti AC, Frauches TS, de Mello CMB, Galliez
420 RM, Faffe DS, Castiñeiras TMPP, Tanuri A, de Vasconcelos ATR: Genomic characterization of a
421 novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. medRxiv 2020:2020.12.23.20248598.
422 [18] Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L: Increased
423 hazard of mortality in cases compatible with SARS-CoV-2 variant of concern 202012/1 - a
424 matched cohort study. medRxiv 2021:2021.02.09.21250937.

[19] Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao M, Korsman S, Davies M-A, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Pond SLK, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T: 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 2020:2020.12.21.20248640.

[20] Bradshaw D, Laubscher R, Dorrington R, Groenewald P, Moultrie T: Report on Weekly Deaths in South Africa: 1 January-8 December 2020 (Week 49). Burden of Disease Research Unit, South African Medical Research Council 2020.

[21] Iacobucci G: Covid-19: New UK variant may be linked to increased death rate, early data indicate. BMJ 2021, 372:n230.

[22] Faria NR, Claro IM, Candido D, Franco LAM, Andrade PS, Coletti TM, Silva CAM, Sales FC, Manuli ER, Aguiar RS, Gaburo N, Camilo CdC, Fraiji NA, Crispim MAE, Carvalho MdPSS, Rambaut A, Loman N, Pybus OG, Sabino EC, Network CG: Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. virological.org, 2021.

[23] Alpert T, Lasek-Nesselquist E, Brito AF, Valesano AL, Rothman J, MacKay MJ, Petrone ME, Breban MI, Watkins AE, Vogels CBF, Russell A, Kelly JP, Shudt M, Plitnick J, Schneider E, Fitzsimmons WJ, Khullar G, Metti J, Dudley JT, Nash M, Wang J, Liu C, Hui P, Muyombwe A, Downing R, Razeq J, Bart SM, Murphy S, Neal C, Laszlo E, Landry ML, Cook PW, Fauver JR, Mason CE, Lauring AS, St George K, MacCannell DR, Grubaugh ND: Early introductions and

community transmission of SARS-CoV-2 variant B.1.1.7 in the United States. medRxiv 2021:2021.02.10.21251540.

[24] Washington NL, Gangavarapu K, Zeller M, Bolze A, Cirulli ET, Schiabor Barrett KM, Larsen BB, Anderson C, White S, Cassens T, Jacobs S, Levan G, Nguyen J, Ramirez JM, Rivera-Garcia C, Sandoval E, Wang X, Wong D, Spencer E, Robles-Sikisaka R, Kurzban E, Hughes LD, Deng X, Wang C, Servellita V, Valentine H, De Hoff P, Seaver P, Sathe S, Gietzen K, Sickler B, Antico J, Hoon K, Liu J, Harding A, Bakhtar O, Basler T, Austin B, Isaksson M, Febbo P, Becker D, Laurent M, McDonald E, Yeo GW, Knight R, Laurent LC, de Feo E, Worobey M, Chiu C, Suchard MA, Lu JT, Lee W, Andersen KG: Genomic epidemiology identifies emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. medRxiv 2021:2021.02.06.21251159.

[25] Jackson A: 90 cases of UK COVID-19 variant B.1.1.7 reported at Michigan prison, state says. Detroit Free Press. Detroit: Detroit Free Press, 2021.

[26] 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, 70:95-9.

[27] B.1.1.7. NaERVTAGNoncC-v: NERVTAG paper on COVID-19 variant of concern B.1.1.7. Edited by Care DoHaS, Emergencies SAGf. United Kingdom: GOV.UK, 2021.

[28] Munitz A, Yechezkel M, Dickstein Y, Yamin D, Gerlic M: The rise of SARS-CoV-2 variant B.1.1.7 in Israel intensifies the role of surveillance and vaccination in elderly. medRxiv 2021:2021.02.16.21251819.

468 [29] Wang WB, Liang Y, Jin YQ, Zhang J, Su JG, Li QM: E484K mutation in SARS-CoV-2 RBD
469 enhances binding affinity with hACE2 but reduces interactions with neutralizing antibodies and
470 nanobodies: Binding free energy calculation studies. bioRxiv 2021.
471 [30] Garcia-Beltran WF, Lam EC, Denis KS, Nitido AD, Garcia ZH, Hauser BM, Feldman J, Pavlovic
472 MN, Gregory DJ, Poznansky MC, Sigal A, Schmidt AG, Iafrate AJ, Naranbhai V, Balazs AB:
473 Circulating SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity.
474 medRxiv 2021.
475 [31] Liu H, Wei P, Zhang Q, Chen Z, Aviszus K, Downing W, Peterson S, Reynoso L, Downey GP,
476 Frankel SK, Kappler J, Marrack P, Zhang G: 501Y.V2 and 501Y.V3 variants of SARS-CoV-2 lose
477 binding to Bamlanivimab in vitro. bioRxiv 2021.
478 [32] Yuan M, Huang D, Lee C-CD, Wu NC, Jackson AM, Zhu X, Liu H, Peng L, van Gils MJ, Sanders
479 RW, Burton DR, Reincke SM, Prüss H, Kreye J, Nemazee D, Ward AB, Wilson IA: Structural and
480 functional ramifications of antigenic drift in recent SARS-CoV-2 variants. bioRxiv 2021.
481 [33] Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD:
482 Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect
483 recognition by polyclonal human plasma antibodies. Cell host & microbe 2021.
484 [34] Tada T, Dcosta BM, Samanovic-Golden M, Herati RS, Cornelius A, Mulligan MJ, Landau NR:
485 Neutralization of viruses with European, South African, and United States SARS-CoV-2 variant
486 spike proteins by convalescent sera and BNT162b2 mRNA vaccine-elicited antibodies. bioRxiv
487 2021.
488 [35] Planas D, Bruel T, Grzelak L, Guivel-Benhassine F, Staropoli I, Porrot F, Planchais C,
489 Buchrieser J, Rajah MM, Bishop E, Albert M, Donati F, Behillil S, Enouf V, Maquart M, Gonzalez

M, De Sèze J, Péré H, Veyer D, Sève A, Simon-Lorière E, Fafi-Kremer S, Stefic K, Mouquet H, Hocqueloux L, van der Werf S, Prazuck T, Schwartz O: Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. bioRxiv 2021.

[36] Wang P, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD, Graham BS, Mascola JR, Chang JY, Yin MT, Sobieszczyk M, Kyratsous CA, Shapiro L, Sheng Z, Nair MS, Huang Y, Ho DD: Increased Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7 to Antibody Neutralization. bioRxiv 2021.

[37] Cele S, Gazy I, Jackson L, Hwa S-H, Tegally H, Lustig G, Giandhari J, Pillay S, Wilkinson E, Naidoo Y, Karim F, Ganga Y, Khan K, Balazs AB, Gosnell BI, Hanekom W, Moosa M-YS, Lessells RJ, de Oliveira T, Sigal A: Escape of SARS-CoV-2 501Y.V2 variants from neutralization by convalescent plasma. medRxiv 2021.

[38] Nelson G, Buzko O, Spilman P, Niazi K, Rabizadeh S, Soon-Shiong P: Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. bioRxiv 2021.

[39] Tian F, Tong B, Sun L, Shi S, Zheng B, Wang Z, Dong X, Zheng P: Mutation N501Y in RBD of Spike Protein Strengthens the Interaction between COVID-19 and its Receptor ACE2. bioRxiv 2021.

[40] Cline M, Emerson M, bratter j, howell j, Jeanty P: Houston Region Grows More Racially/Ethnically Diverse,With Small Declines in Segregation.A Joint Report Analyzing Census Data from 1990, 2000, and 2010, 2012.

511 [41] Wright AM, Beres SB, Consamus EN, Long SW, Flores AR, Barrios R, Richter GS, Oh SY,
512 Garufi G, Maier H, Drews AL, Stockbauer KE, Cernoch P, Schneewind O, Olsen RJ, Musser JM:
513 Rapidly progressive, fatal, inhalation anthrax-like infection in a human: case report, pathogen
514 genome sequencing, pathology, and coordinated response. Arch Pathol Lab Med 2011,
515 135:1447-59.
516 [42] Long SW, Beres SB, Olsen RJ, Musser JM: Absence of patient-to-patient intrahospital
517 transmission of *Staphylococcus aureus* as determined by whole-genome sequencing. MBio
518 2014, 5:e01692-14.
519 [43] Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis JJ, Brettin T, Xia F, Musser JM:
520 Population Genomic Analysis of 1,777 Extended-Spectrum Beta-Lactamase-Producing Klebsiella
521 pneumoniae Isolates, Houston, Texas: Unexpected Abundance of Clonal Group 307. MBio 2017,
522 8.
523 [44] Nasser W, Beres SB, Olsen RJ, Dean MA, Rice KA, Long SW, Kristinsson KG, Gottfredsson M,
524 Vuopio J, Raisanen K, Caugant DA, Steinbakk M, Low DE, McGeer A, Darenberg J, Henriques-
525 Normark B, Van Beneden CA, Hoffmann S, Musser JM: Evolutionary pathway to increased
526 virulence and epidemic group A *Streptococcus* disease derived from 3,615 genome sequences.
527 Proc Natl Acad Sci U S A 2014, 111:E1768-76.
528 [45] Kachroo P, Eraso JM, Beres SB, Olsen RJ, Zhu L, Nasser W, Bernard PE, Cantu CC, Saavedra
529 MO, Arredondo MJ, Strope B, Do H, Kumaraswami M, Vuopio J, Grondahl-Yli-Hannuksela K,
530 Kristinsson KG, Gottfredsson M, Pesonen M, Pensar J, Davenport ER, Clark AG, Corander J,
531 Caugant DA, Gaini S, Magnussen MD, Kubiak SL, Nguyen HAT, Long SW, Porter AR, DeLeo FR,

532 Musser JM: Integrated analysis of population genomics, transcriptomics and virulence provides
533 novel insights into *Streptococcus pyogenes* pathogenesis. Nat Genet 2019, 51:548-59.
534 [46] Long SW, Olsen RJ, Christensen PA, Bernard DW, Davis JJ, Shukla M, Nguyen M, Saavedra
535 MO, Yerramilli P, Pruitt L, Subedi S, Kuo HC, Hendrickson H, Eskandari G, Nguyen HAT, Long JH,
536 Kumaraswami M, Goike J, Boutz D, Gollihar J, McLellan JS, Chou CW, Javanmardi K, Finkelstein
537 IJ, Musser JM: Molecular Architecture of Early Dissemination and Massive Second Wave of the
538 SARS-CoV-2 Virus in a Major Metropolitan Area. mBio 2020, 11.
539 [47] Salazar E, Kuchipudi SV, Christensen PA, Eagar T, Yi X, Zhao P, Jin Z, Long SW, Olsen RJ,
540 Chen J, Castillo B, Leveque C, Towers D, Lavinder J, Gollihar J, Cardona J, Ippolito G, Nissly R,
541 Bird I, Greenawalt D, Rossi RM, Gontu A, Srinivasan S, Poojary I, Cattadori IM, Hudson PJ,
542 Josleyn NM, Prugar L, Huie K, Herbert A, Bernard DW, Dye JM, Kapur V, Musser JM:
543 Convalescent plasma anti-SARS-CoV-2 spike protein ectodomain and receptor-binding domain
544 IgG correlate with virus neutralization. J Clin Invest 2020, 130:6728-38.
545 [48] Salazar E, Perez KK, Ashraf M, Chen J, Castillo B, Christensen PA, Eubank T, Bernard DW,
546 Eagar TN, Long SW, Subedi S, Olsen RJ, Leveque C, Schwartz MR, Dey M, Chavez-East C, Rogers
547 J, Shehabeldin A, Joseph D, Williams G, Thomas K, Masud F, Talley C, Dlouhy KG, Lopez BV,
548 Hampton C, Lavinder J, Gollihar JD, Maranhao AC, Ippolito GC, Saavedra MO, Cantu CC,
549 Yerramilli P, Pruitt L, Musser JM: Treatment of Coronavirus Disease 2019 (COVID-19) Patients
550 with Convalescent Plasma. Am J Pathol 2020, 190:1680-90.
551 [49] Long SW, Olsen RJ, Christensen PA, Bernard DW, Davis JR, Shukla M, Nguyen M, Ojeda
552 Saavedra M, Cantu CC, Yerramilli P, Pruitt L, Subedi S, Hendrickson H, Eskandari G,
553 Kumaraswami M, McLellan JS, Musser JM: Molecular Architecture of Early Dissemination and

554 Evolution of the SARS-CoV-2 Virus in Metropolitan Houston, Texas. bioRxiv
555 2020:2020.05.01.072652.
556 [50] Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N,
557 Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C,
558 Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R,
559 Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D,
560 Yoo H, Stevens R: The PATRIC Bioinformatics Resource Center: expanding data and analysis
561 capabilities. Nucleic Acids Res 2020, 48:D606-D12.
562 [51] Li H: Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018,
563 34:3094-100.
564 [52] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
565 Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools.
566 Bioinformatics 2009, 25:2078-9.
567 [53] Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, Tan AL, Paul LM,
568 Brackney DE, Grewal S, Gurfield N, Van Rompay KKA, Isern S, Michael SF, Coffey LL, Loman NJ,
569 Andersen KG: An amplicon-based sequencing framework for accurately measuring intrahost
570 virus diversity using PrimalSeq and iVar. Genome Biol 2019, 20:8.
571 [54] Shu Y, McCauley J: GISAID: Global initiative on sharing all influenza data - from vision to
572 reality. Euro Surveill 2017, 22.
573 [55] COVID-19 Genomics UK Consortium. 2021.

574 [56] Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH: Increased hazard of
575 death in community-tested cases of SARS-CoV-2 Variant of Concern 202012/01. medRxiv
576 2021:2021.02.01.21250959.
577 [57] Muik A, Wallisch AK, Sanger B, Swanson KA, Muhl J, Chen W, Cai H, Maurus D, Sarkar R,
578 Tureci O, Dormitzer PR, Sahin U: Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by
579 BNT162b2 vaccine-elicited human sera. Science 2021:eabg6105.
580 [58] Starr TN, Greaney AJ, Dingens AS, Bloom JD: Complete map of SARS-CoV-2 RBD mutations
581 that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. bioRxiv
582 2021:2021.02.17.431683.
583 [59] Sabino EC, Buss LF, Carvalho MPS, Prete CA, Jr., Crispim MAE, Fraiji NA, Pereira RHM, Parag
584 KV, da Silva Peixoto P, Kraemer MUG, Oikawa MK, Salomon T, Cucunuba ZM, Castro MC, de
585 Souza Santos AA, Nascimento VH, Pereira HS, Ferguson NM, Pybus OG, Kucharski A, Busch MP,
586 Dye C, Faria NR: Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. Lancet
587 2021, 397:452-5.
588 [60] Latif AA, Gangavarapu K, Haag E, Matteson N, Mullen JL, Tsueng G, Zeller M, Wu C, Su AI,
589 Hughes LD, Andersen KG, Biology CfVS: P.2 Lineage Report. outbreak.info, 2021.
590 [61] O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone J, Ruis C, Abu-Dahab K, Taylor
591 B, Yeats C, Plessis Ld, Aanensen D, Holmes E, Pybus O, Rambaut A: pangolin: lineage assignment
592 in an emerging pandemic as an epidemiological tool. 2021.
593 [62] Latif AA, Gangavarapu K, Haag E, Matteson N, Mullen JL, Tsueng G, Zeller M, Wu C, Su AI,
594 Hughes LD, Andersen KG, Biology CfVS: B.1.427 Lineage Report. 2021.

595 [63] Latif AA, Gangavarapu K, Haag E, Matteson N, Mullen JL, Tsueng G, Zeller M, Wu C, Su AI,
596 Hughes LD, Andersen KG, Biology CfVS: B.1.429 Lineage Report. outbreak.info, 2021.
597 [64] Grabowski F, Kochańczyk M, Lipniacki T: L18F substrain of SARS-CoV-2 VOC-202012/01 is
598 rapidly spreading in England. medRxiv 2021.
599 [65] DeWitt M: Rapid Impact Analysis of B 1.1.7 Variant on the Spread of SARS-CoV-2 in North
600 Carolina. medRxiv 2021.
601 [66] Younes M, Hamze K, Nassar H, Makki M, Ghadar M, Nguewa P, Sater FA: Emergence and
602 fast spread of B.1.1.7 lineage in Lebanon. medRxiv 2021.
603 [67] Vasques Nonaka CK, Miranda Franco M, Gräf T, Almeida Mendes AV, Santana de Aguiar R,
604 Giovanetti M, Solano de Freitas Souza B: Genomic Evidence of a Sars-Cov-2 Reinfection Case
605 With E484K Spike Mutation in Brazil. preprintsorg 2021.
606 [68] Resende PC, Bezerra JF, Vasconcelos RHTd, Ighor Arantes3 LA, Mendonça AC, Paixao AC,
607 Rodrigues ACD, Silva T, Rocha AS, Pauvolid-Corrêa A, Motta FC, Teixeira DLF, Carneiro TFdO,
608 Neto FPF, Herbster ID, Leite AB, Riediger IN, Debur MdC, Naveca FG, Almeida W, Livorati M,
609 Bello G, Siqueira MM: Spike E484K mutation in the first SARS-CoV-2 reinfection case confirmed
610 in Brazil. 2021.
611 [69] Wise J: Covid-19: The E484K mutation and the risks it poses. BMJ 2021, 372:n359.
612 [70] Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM, Theel
613 ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH, Diamond MS, Whelan SPJ:
614 Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody
615 neutralization. Cell host & microbe 2021.

616 [71] Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, Schaefer-Babajew D,
617 Cipolla M, Gaebler C, Lieberman JA, Oliveira TY, Yang Z, Abernathy ME, Huey-Tubman KE,
618 Hurley A, Turroja M, West KA, Gordon K, Millard KG, Ramos V, Silva JD, Xu J, Colbert RA, Patel R,
619 Dizon J, Unson-O'Brien C, Shimeliovich I, Gazumyan A, Caskey M, Bjorkman PJ, Casellas R,
620 Hatziioannou T, Bieniasz PD, Nussenzweig MC: mRNA vaccine-elicited antibodies to SARS-CoV-2
621 and circulating variants. Nature 2021.
622 [72] Pater AA, Bosmeny MS, Barkau CL, Ovington KN, Chilamkurthy R, Parasrampuria M,
623 Eddington SB, Yinusa AO, White AA, Metz PE, Sylvain RJ, Hebert MM, Benzinger SW, Sinha K,
624 Gagnon KT: Emergence and Evolution of a Prevalent New SARS-CoV-2 Variant in the United
625 States. bioRxiv 2021:2021.01.11.426287.
626 [73] Hodcroft EB, Domman DB, Snyder DJ, Oguntuyo K, Van Diest M, Densmore KH, Schwalm
627 KC, Femling J, Carroll JL, Scott RS, Whyte MM, Edwards MD, Hull NC, Kevil CG, Vanchiere JA, Lee
628 B, Dinwiddie DL, Cooper VS, Kamil JP: Emergence in late 2020 of multiple lineages of SARS-CoV-
629 2 Spike protein variants affecting amino acid position 677. medRxiv 2021:2021.02.12.21251658.
630 [74] COVID-19 Positive Cumulative Cases. Texas Medical Center COVID-19 Dashboard: TMC,
631 2021.
632 [75] Vavrek D, Speroni L, Curnow KJ, Oberholzer M, Moeder V, Febbo PG: Genomic surveillance
633 at scale is required to detect newly emerging strains at an early timepoint. medRxiv
634 2021:2021.01.12.21249613.
635 [76] Burki T: Understanding variants of SARS-CoV-2. Lancet 2021, 397:462.
636 [77] White House Briefing Room: President Biden Announces American Rescue Plan. 2021.
637

638

639

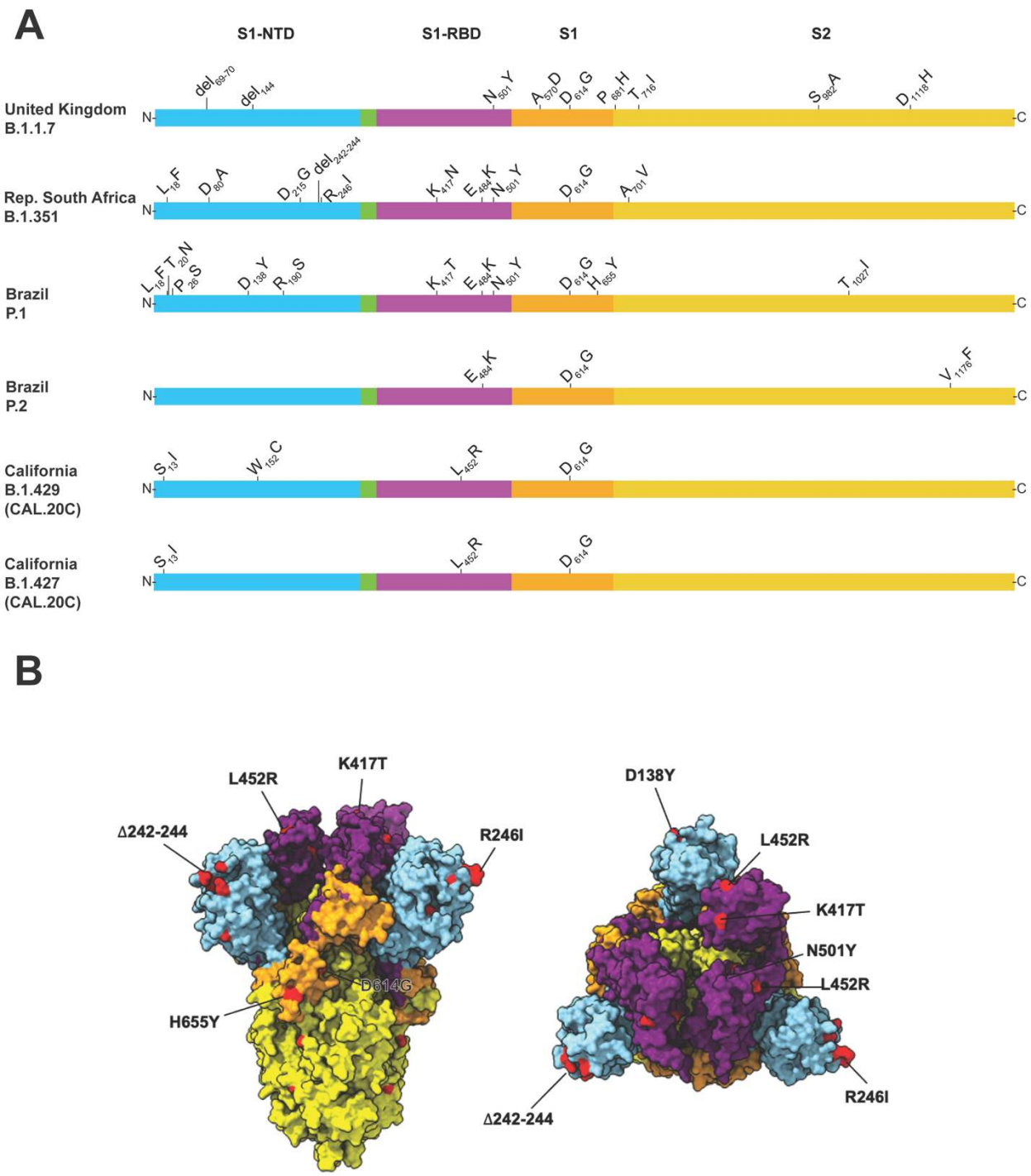
640 **Table 1. Variants of concern or variant of interest identified in the Houston Metropolitan**

641 **area.**

Variant	No. of Isolates
B.1.1.7	23
B.1.351	2
P.1	4
P.2	39
B.1.429	143
B.1.427	19

642

643



644
645 **Figure 1. A:** Schematic showing structural changes present in the spike protein of the major

646 SARS.CoV.2 variants identified in the study. S1-NTD, S1 domain-aminoterminal domain; S1-RBD,

647 S1 domain-receptor binding domain; S1, S1 domain; S2, S2 domain. **B:** Mapping of important

changes onto the cryoEM structure of spike protein. The color scheme matches that used in panel A. Blue (NTD), purple (RBD), orange (S1), and yellow (S2). Aggregate mutations present in variants of concern are colored in red when amino acid residues are present in the resolved structure. Left, side view of SARS-CoV-2 prefusion-stabilized spike. Right, top view. Structure of PDB 6vsb was used as reference.

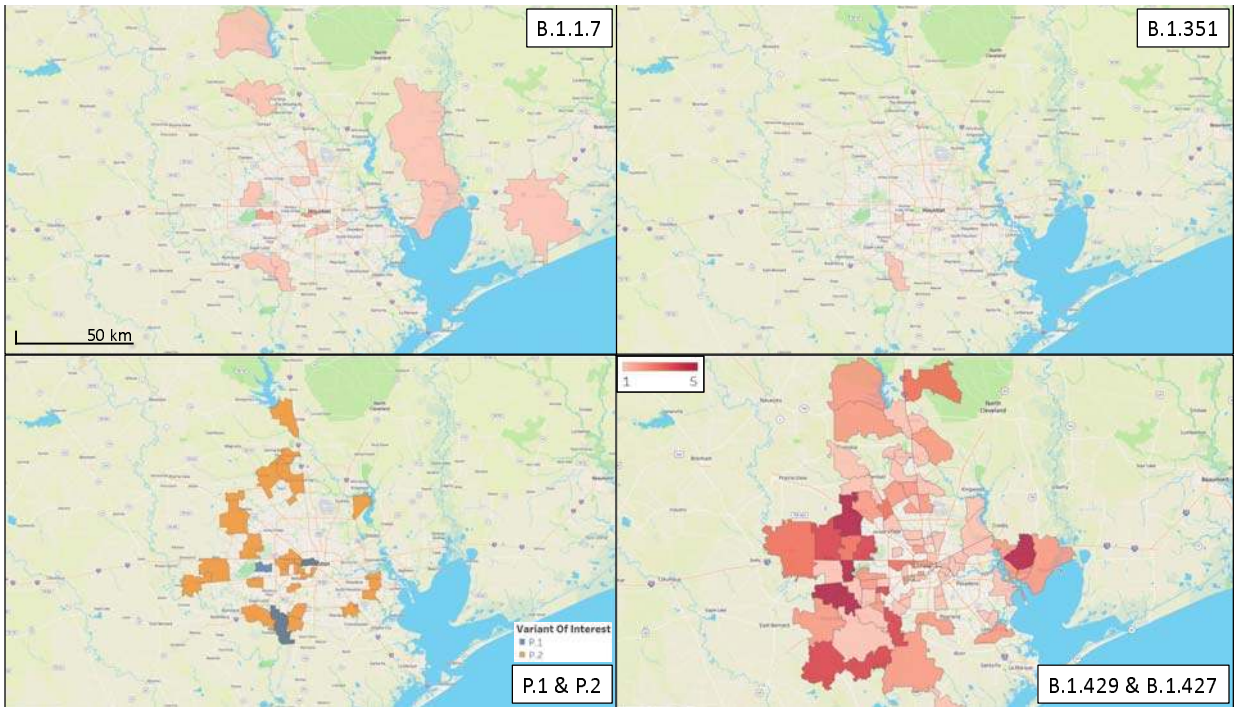


Figure 2. Geospatial distribution for each variant of concern identified in the study. The home address zip code for each patient was used and figures were generated using Tableau version 2020.3.4.