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 InfecnNEB25NionrF

Number of te t pages: 12

Number of tables: 1

Number of figures: 2

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

[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 e tensive 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-] 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).[1 -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 e ample, the UK B.1.1.7 variant spread rapidly in southeast England where it caused large numbers of COVID-19 cases,[1] and was identified shortly thereafter in the United States (US)

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 itro* 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 appro imately 7 million (https://www.houston.org/houston-data). [0] 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. [1- 9] 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. [6, 9]

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. [6, 9] We have continued to sequence positive SARS-CoV-2 specimens with the goal of monitoring for variants of concern and genome mutations

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

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)

196 is associated with increased transmissibility[26]. In addition, evidence has been presented from
197 the UK that B.1.1.7 strains may cause increased hospitalization and mortality.[18, 21, 27, 56] The first
198 patient we identified in Houston with a B.1.1.7 variant was diagnosed the second week of
199 January, 2020; thus far we have identified 23 patients with this variant of concern (Table 1). Of
200 note, none of our first three patients had an international travel history, suggesting that they
201 acquired the B.1.1.7 infections either locally or during domestic travel. Preliminary evidence
202 indicates that immune sera from the Pfizer-BioNTech SARS-CoV-2 vaccine retain the ability to
203 neutralize B.1.1.7 variants *in itro*.[57] Additional studies have found that convalescent plasma
204 from many patients, and some monoclonal antibody therapies, retain the ability to neutralize
205 B.1.1.7 variant SARS-CoV-2 *in itro*.[3, 35]
206
207 ## South Africa Variant of Concern (B.1.351)
208 The South Africa B.1.351 variant of concern was first identified in a COVID-19 epidemic wave
209 occurring in Nelson Mandela Bay in October 2020.[19] This variant was concerning because of its
210 large number of spike protein mutations (including K417N, E484K, and N501Y) (Figure 1) and
211 apparent increased transmissibility.[19, 38] These three mutations are located in the receptor
212 binding domain of spike and may decrease the effectiveness of some mAb therapies and
213 vaccines.[29-31, 3, 35, 58] The first South Africa variant detected in Houston was identified in a

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 e amined their geospatial distribution to investigate the e tent of dissemination (Figure 2). With the e ception of the B.1.351 variant, patients infected with all

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 une pected but it is disquieting. With this report, Houston now

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 si 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 e tensive 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.[5] Since March 2020, appro imately
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,[7] 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 metrople . 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 machine7 u3-8 1K389s6l7f5Ktfble359 o 9G9fve 6l7f5 de patsèKcosG89ch

327

328

329

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.,

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 Ta onomy 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.

382 [7] Wang C, Horby PW, Hayden FG, Gao GF: A novel coronavirus outbreak of global health
383 concern. Lancet 2020, 395:470-3.
384 [8] Perlman S: Another Decade, Another Coronavirus. New England Journal of Medicine 2020,
385 382:760-2.
386 [9] Allel K, Tapia-Muñoz T, Morris W: Country-level factors associated with the early spread of
387 COVID-19 cases at 5, 10 and 15 days since the onset. Glob Public Health 2020:1-14.
388 [10] Chen P, Nirula A, Heller B, Gottlieb RL, Boscia J, Morris J, Huhn G, Cardona J, Mocherla B,
389 Stosor V, Shawa I, Adams AC, Van Naarden J, Custer KL, Shen L, Durante M, Oakley G, Schade
390 AE, Sabo J, Patel DR, Klekotka P, Skovronsky DM, Investigators B-: SARS-CoV-2 Neutralizing
391 Antibody LY-CoV555 in Outpatients with Covid-19. N Engl J Med 2021, 384:229-37.
392 [11] Weinreich DM, Sivapalasingam S, Norton T, Ali S, Gao H, Bhore R, Musser BJ, Soo Y, Rofail
393 D, Im J, Perry C, Pan C, Hosain R, Mahmood A, Davis JD, Turner KC, Hooper AT, Hamilton JD,
394 Baum A, Kyratsous CA, Kim Y, Cook A, Kampman W, Kohli A, Sachdeva Y, Graber X, Kowal B,
395 DiCioccio T, Stahl N, Lipsich L, Braunstein N, Herman G, Yancopoulos GD, Trial I: REGN-COV2, a
396 Neutralizing Antibody Cocktail, in Outpatients with Covid-19. N Engl J Med 2021, 384:238-51.
397 [12] Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, Diemert D, Spector SA,
398 Rouphael N, Creech CB, McGettigan J, Khetan S, Segall N, Solis J, Brosz A, Fierro C, Schwartz H,
399 Neuzil K, Corey L, Gilbert P, Janes H, Follmann D, Marovich M, Mascola J, Polakowski L,
400 Ledgerwood J, Graham BS, Bennett H, Pajon R, Knightly C, Leav B, Deng W, Zhou H, Han S,
401 Ivarsson M, Miller J, Zaks T, Group CS: Efficacy and Safety of the mRNA-1273 SARS-CoV-2
402 Vaccine. N Engl J Med 2021, 384:403-16.

[13] Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Perez Marc G, Moreira ED, Zerbini C, Bailey R, Swanson KA, Roychoudhury S, Koury K, Li P, Kalina WV, Cooper D, Frenck RW, Jr., Hammitt LL, Tureci O, Nell H, Schaefer A, Unal S, Tresnan DB, Mather S, A, Ls Dormitzer PR, mRNA Covid-19 Vaccine. N Engl J Med 2020, 383:2603-1 .

[14] Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, Connor T, Peacock T,

[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. medR iv 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,

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. bioR iv 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 medR iv 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
binding to BamyF)5K] GGFv 2:d 5670F)5K01Y670F)5K.V3 8J,88S-670F)5K2vG86ñs9F9nt22vG82vG8.V17:t2:ivG8

M, De Sèze J, Péré H, Veyer D, Sève A, Simon-Lorière E, Fafi-Kremer S, Stefic K, Mouquet H, Hocquelou 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. bioR iv 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. bioR iv 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. medR iv 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. bioR iv 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. bioR iv 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:

Musser JM: Integrated analysis of population genomics, transcriptomics and virulence provides novel insights into *Streptococcus pyogenes* pathogenesis. Nat Genet 2019, 51:548-59.

[46] Long SW, Olsen RJ, Christensen PA, Bernard DW, Davis JJ, Shukla M, Nguyen M, Saavedra MO, Yerramilli P, Pruitt L, Subedi S, Kuo HC, Hendrickson H, Eskandari G, Nguyen HAT, Long JH, Kumaraswami M, Goike J, Boutz D, Gollihar J, McLellan JS, Chou CW, Javanmardi K, Finkelstein IJ, Musser JM: Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area. mBio 2020, 11.

[47] Salazar E, Kuchipudi SV, Christensen PA, Eagar T, Yi X, Zhao P, Jin Z, Long SW, Olsen RJ,

554 Evolution of the SARS-CoV-2 Virus in Metropolitan Houston, Te as. bioR iv

555 2020:2020.05.01.072652.

5551 5551 a os09GtlGGGGG8SAG07R,G7989 os09G7989 os8sG a4

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. medR iv
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.

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. medR iv 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. medR iv 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. medR iv 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, Pai ao AC,
607 Rodrigues ACD, Silva T, Rocha AS, Pauvolid-Corrêa A, Motta FC, Tei eira 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.

[71] Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, Schaefer-Babajew D, Cipolla M, Gaebler C, Lieberman JA, Oliveira TY, Yang Z, Abernathy ME, Huey-Tubman KE, Hurley A, Turroja M, West KA, Gordon K, Millard KG, Ramos V, Silva JD, Xu J, Colbert RA, Patel R, Dizon J, Unson-O'Brien C, Shimeliovich I, Gazumyan A, Caskey M, Bjorkman PJ, Casellas R, Hatziioannou T, Bieniasz PD, Nussenzweig MC: mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature 2021.

[72] Pater AA, Bosmeny MS, Barkau CL, Ovington KN, Chilamkurthy R, Parasrampuria M, Eddington SB, Yinusa AO, White AA, Metz PE, Sylvain RJ, Hebert MM, Benzinger SW, Sinha K, Gagnon KT: Emergence and Evolution of a Prevalent New SARS-CoV-2 Variant in the United States. bioR iv 2021:2021.01.11.426287.

[73] Hodcroft EB, Domman DB, Snyder DJ, Oguntuyo K, Van Diest M, Densmore KH, Schwalm KC, Femling J, Carroll JL, Scott RS, Whyte MM, Edwards MD, Hull NC, Kevil CG, Vanchiere JA, Lee B, Dinwiddie DL, Cooper VS, Kamil JP: Emergence in late 2020 of multiple lineages of SARS-CoV-2 Spike protein variants affecting amino acid position 677. medR iv 2021:2021.02.12.21251658.

[74] COVID-19 Positive Cumulative Cases. Te as Medical Center COVID-19 Dashboard: TMC, 2021.

[75] Vavrek D, Speroni L, Curnow KJ, Oberholzer M, Moeder V, Febbo PG: Genomic surveillance at scale is required to detect newly emerging strains at an early timepoint. medR iv 2021:2021.01.12.21249613.

[76] Burki T: Understanding variants of SARS-CoV-2. Lancet 2021, 397:462.

[77] White House Briefing Room: President Biden Announces American Rescue Plan. 2021.

638

639

Table 1. Variants of on ern or variant of interest identified in the Houston etropolitan 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

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 647S1 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.

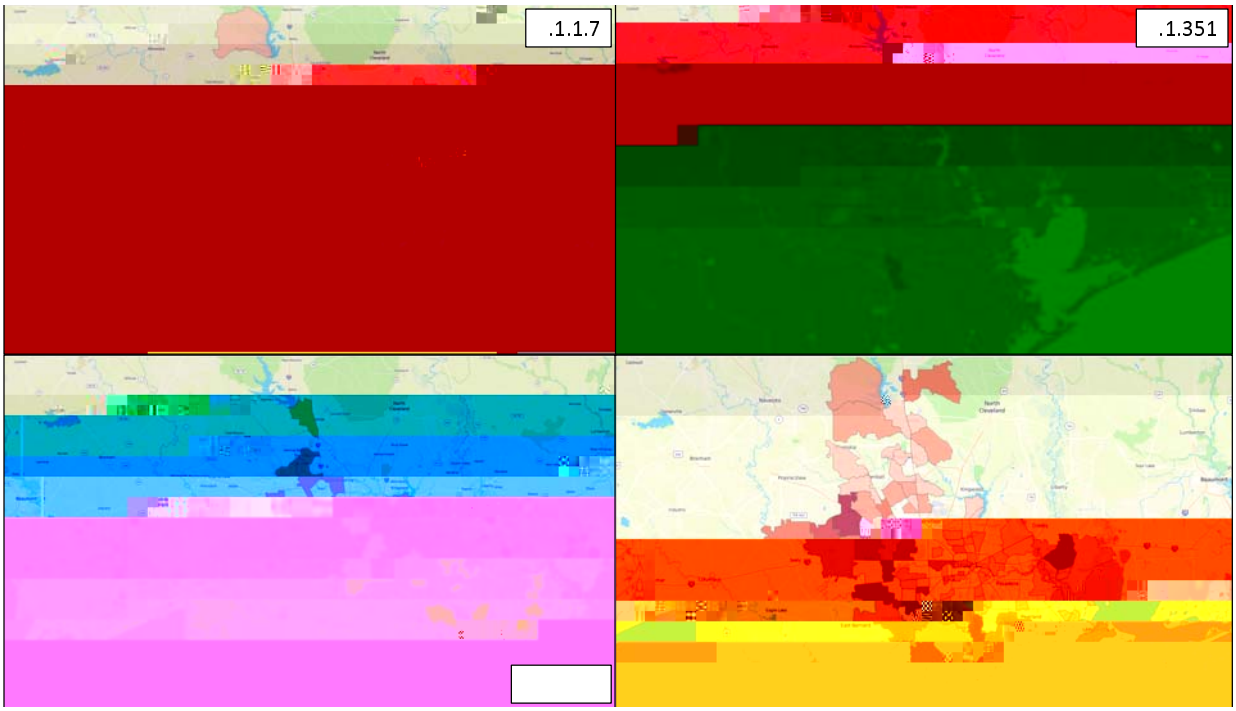

Figure . Geospatial distribution for each variant of concern identified in the study. The home address zip code for each patient was used and figures were generated using Tableau version 2020.3.4.