

1 **Emergence of the E484K Mutation in SARS-CoV-2 Lineage B.1.1.220 in Upstate New York**

2 Emil Lesho<sup>1\*#</sup>, Brendan Corey<sup>2\*</sup>, Francois Lebreton<sup>2</sup>, Ana C. Ong<sup>2</sup>, Brett E. Swierczewski<sup>2</sup>,  
3 Jason W. Bennett<sup>2</sup>, Edward E. Walsh<sup>1,3</sup>, and Patrick Mc Gann<sup>2#</sup>

4

5 <sup>1</sup> Infectious Diseases Unit, Rochester General Hospital, Rochester, New York, USA

6 <sup>2</sup> Multidrug-Resistant Organism Repository and Surveillance Network (MRSN), Walter Reed

7 Army Institute of Research, Silver Spring, Maryland, USA

8 <sup>3</sup> University of Rochester School of Medicine, Rochester, NY

9

10 \* Emil Lesho and Brendan Corey contributed equally to this article

11

12 #Corresponding author

# Alternate corresponding author

13 Patrick Mc Gann, PhD

Emil P. Lesho, DO

14 Walter Reed Army Institute of Research

Infectious Diseases Unit

15 503 Robert Grant Avenue, 2A36

Rochester General Hospital

16 Silver Spring, MD 20910, USA

Rochester, NY 14850, USA

17 Email: [patrick.t.mcgann4.civ@mail.mil](mailto:patrick.t.mcgann4.civ@mail.mil)

Email: [emil.lesho@rochesterregional.org](mailto:emil.lesho@rochesterregional.org)

18 **Abstract**

19 Ongoing surveillance detected a SARS-CoV-2 B.1.1.220 variant carrying the E484K substitution  
20 in four patients from a hospital network in upstate New York. Patients reported no travel history  
21 and shared no obvious epidemiological linkage. A search of online databases identified 12  
22 additional B.1.1.220 with E484K, all of which were detected in New York since December 2020.  
23 Detailed genomic analyses suggests that the mutation has emerged independently in at least two  
24 different B.1.1.220 strains in this region.

25 **Text**

26 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the etiological agent of  
27 coronavirus disease 2019 (COVID-19), has caused a global pandemic that is reshaping society.  
28 As part of an unprecedented global response from the scientific community the development and  
29 rapid deployment of multiple antibody-based countermeasures that target the viral spike protein  
30 has proceeded at an incredible pace.

31 Over the past six months, the emergence of three variants with mutations in the spike  
32 protein has raised serious concerns about the durability of the current suite of vaccines and  
33 immunotherapies. The variants, which are colloquially referred to as the United Kingdom  
34 (B.1.1.7 using the Pangolin nomenclature<sup>1</sup>), South African (B.1.351), and Brazilian (P.1)  
35 variants<sup>2-4</sup>, have been increasingly identified worldwide<sup>5</sup>. The first variant (B.1.1.7) was  
36 identified in September 2020 and carries a N501Y mutation in the receptor-binding domain  
37 (RBD) of the spike protein<sup>4</sup>. Critically, recent studies have suggested this mutation potentially  
38 increases transmission and virulence<sup>6,7</sup>. Shortly thereafter, this same mutation was identified in  
39 B.1.351 and P.1<sup>2,3</sup>. Both variants also carried an additional substitution (E484K) in the spike that  
40 can increase resistance to neutralization by many monoclonal antibodies (mAb), while most  
41 convalescent sera and mRNA vaccine-induced immune sera show reduced inhibitory activity<sup>8-11</sup>.  
42 Notably, interim results from the Novavax adjuvanted spike protein nanoparticle COVID-19  
43 vaccine trial indicated an efficacy of 85.6% against B.1.1.7 infection (versus 95.6% for the  
44 original strain), but just 49.4% to 60% in a South Africa Phase 2b trial in which the B.1.351  
45 strain was detected in the majority of COVID-19 events<sup>12</sup>.

46 In the United States, there have been 3,037 cases with the B.1.1.7 variant identified in 49  
47 jurisdictions, 81 cases with B.1.351 (20 jurisdictions), and 15 cases with P.1 (9 jurisdictions) as

48 of this writing (March 2021)<sup>13</sup>. In addition, several regional variants have been identified, such  
49 as the 20C-US lineage in the Midwest<sup>14</sup> and the B.1.2 lineage carrying the Q677P substitution in  
50 New Mexico and Louisiana<sup>15</sup>. However, these lack the E484K and N501Y mutations. In  
51 contrast, the recently described B.1.536 variant in New York City carries E484K<sup>16</sup>. Notably, this  
52 variant has been spreading at an alarming rate since November and by February 15<sup>th</sup> 2021  
53 accounted for 12.3% of all cases in this catchment area. Similar to previous studies with variants  
54 carrying E484K, antibody neutralization by mAbs REGN10933, CB6, and LY-CoV555 are  
55 either impaired or abolished with this variant<sup>16</sup>. Similarly, the neutralizing activities of  
56 convalescent plasma or post-vaccination sera were lower by 7.7-fold and 3.4-fold, respectively<sup>16</sup>.  
57 These data illustrate the alarming rise of this mutation, which has now been detected in 88 of 886  
58 lineages (as of this writing), with various degrees of frequency<sup>5,17</sup>.

59 In June 2020, a pilot study to investigate SARS-CoV-2 was initiated between the Multi-  
60 drug resistant organism Repository and Surveillance Network (MRSN) and the Rochester  
61 Regional Health System; a network consisting of 8 acute care hospitals, 9 urgent care centers,  
62 and 6 long-term care facilities in St Lawrence County and the 9-county Finger Lakes Region of  
63 NY. Every hospital-based laboratory in the network is capable of testing for COVID-19 infection  
64 and to date 438,107 samples have been tested with 25,468 distinct patients testing positive. For  
65 this study, samples were provided by five hospitals in the Finger Lakes region. RNA is extracted  
66 from SARS-CoV-2-positive nasopharyngeal swabs and transferred to the MRSN for sequencing  
67 (See Methods). To date, >150 samples spanning April 2020 to February 2021 have been  
68 analyzed, with 14 different Pangolin lineages identified (manuscript in preparation). Though  
69 notable mutations were found in the spike proteins from these earlier samples, including D614G,  
70 S673T, Q677H, N679K, and P681H, the E484K and N501Y substitutions were not identified.

71 The most recent shipment consisted of 20 samples from 19 patients collected between  
72 January 27<sup>th</sup> and February 7<sup>th</sup>, 2021. Upon analysis, the E484K substitution was detected in five  
73 samples from one male and three female patients. The average age was 85 years (range 74-92)  
74 and, despite advanced age and multiple co-morbidities (mean CALL score = 10)<sup>18</sup>, all recovered.  
75 Three patients had mild courses with no specific anti-SARS-CoV-2 treatments. One patient  
76 received Remdesivir and Dexamethasone and did not require intensive care or endotracheal  
77 intubation. Notably, three patients lived in separate assisted living facilities with no known  
78 connections and the fourth resided at home. All patients reported no significant travel in the  
79 preceding 6 months.

80 The five strains were assigned to Pangolin Lineage B.1.1.220 (NextStrain clade 20B;  
81 **Figure 1A**), a globally distributed lineage that has also been identified in 2-7% of COVID-19  
82 cases in the North-East US (**Figure 1B**). In addition to the E484K substitution, the strains carried  
83 the characteristic 4 amino acid substitutions that define this lineage when compared to the  
84 Wuhan-Hu-1 SARS-CoV-2 reference genome (NCBI GenBank Acc. MN908947), namely  
85 R203K, G204R, P314L and D614G. This latter substitution has become the most prevalent  
86 mutation world-wide and studies suggest it increases viral transmissibility and results in higher  
87 viral loads<sup>19,20</sup>. To better understand the relationship between the 5 samples we performed a  
88 whole genome high-resolution single nucleotide polymorphism (SNP) based comparison. The  
89 two samples from the same patient were genetically identical (0 SNPs), but the remaining  
90 samples were separated by 1-4 SNPs.

91 Analysis of the five variants in the open data software at MicroReact<sup>21</sup> identified 6  
92 additional B.1.1.220 variants with the E484K substitution. All were deposited by the Wadsworth  
93 Center (Albany, NY), a sentinel institute of the New York Department of Health. Further

94 comparative analysis with 124 B.1.1.220 sequences from the New York region deposited at  
95 GISAID identified an additional 6 variants with this mutation. Surprisingly, a phylogenetic  
96 analysis suggested that the E484K substitution has arisen independently at least twice within  
97 Lineage B.1.1.220 in the New York region, as evidenced by the significant genetic separation  
98 (25 to 29 SNPs) between a monophyletic cluster of 16 samples from Central and Western New  
99 York and sample 1098129 from the New York Metropolitan area (**Figure 1C, Table S1**).  
100 Zoomed-in analysis of the 16 monophyletic strains revealed high-genetic relatedness (0-7 SNPs)  
101 and further suggested that this B.1.1.220 E484K variant has just recently emerged and is rapidly  
102 spreading across New York State with isolates being collected from at least 6 distinct counties  
103 between December and February 2020 (**Figure 1D**). Importantly, as of this writing (March 2021)  
104 the only B.1.1.220 variants shown to harbor the E484K mutation have been isolated in New  
105 York.

106 To the best of our knowledge, this is the first report of a Lineage B.1.1.220 bearing the  
107 important E484K mutation. The observed mutational convergence within the relatively small  
108 subset of Lineage B.1.1.220 genomes from New York State further indicates that codon S/484 is  
109 evolving under a strong degree of positive selection. Though not supported statistically, it is  
110 notable that no B.1.1.220 variants were detected in our surveillance effort prior to December  
111 2020 when it suddenly comprised 25 % of the late January and early February samples.  
112 However, as of this writing (March 2021), only 146 B.1.1.220 genomes from NY have been  
113 uploaded to GISAID, with the majority (n = 87) collected since January 1<sup>st</sup>, 2021; additional  
114 B.1.1.220 sequences from this region are required to better understand the current distribution  
115 and apparent emergence of this concerning variant. It will also be important to assess how the  
116 E484K mutation affects protection by vaccines against this specific variant that lacks the other

117 mutations in the B1.351 and P1 strains. As the vaccine roll-out gathers pace across the United  
118 States, it is imperative that every means necessary are employed to rapidly identify the  
119 emergence of variants-of-concern and immediate action taken to limit their spread.

## 120 **Methods**

### 121 **Sequencing of SARS-CoV-2 samples**

122 Viral RNA was extracted from nasopharyngeal swabs using the QIAamp Viral RNA Mini Kit  
123 (Qiagen, 52906). RNA was converted to cDNA and enriched for SARS-CoV-2 using the QIAseq  
124 SARS-CoV-2 Primer Panel (Qiagen, 333896). Libraries were constructed using the KAPA  
125 HyperPlus kit (Roche, KK8514), quantified with the KAPA Library Quantification kit (Roche,  
126 KK44), pooled, denatured and diluted to 14 pM. The diluted sample was sequenced on an  
127 Illumina MiSeq sequencer using a MiSeq v3 (600 cycle) Reagent Kit (Illumina, MS-102-3003).

### 128 **Bioinformatic Analysis of SARS-CoV-2 genomes**

129 Miseq read data was trimmed for adapters and low quality sequence using BBduk v38.79  
130 (<https://sourceforge.net/projects/bbmap/>). Trimmed reads were then mapped to the Wuhan-Hu-1  
131 SARS-CoV-2 reference genome (NCBI GenBank Acc. MN908947) using Snippy v4.4.5  
132 (<https://github.com/tseemann/snippy>). BAM files were imported into Geneious<sup>22</sup> where variants  
133 identified by Snippy were manually curated/confirmed. Nucleotide alignments were performed  
134 with Muscle<sup>23</sup> followed by trimming of ambiguous/under-represented sequence at the 5'/3' ends  
135 of the alignment in Geneious. Genome-based, maximum likelihood phylogenies were produced  
136 with RAxML-ng (model GTR+Gamma, 50 parsimony, 50 random)<sup>24</sup>. Best scoring phylogenetic  
137 trees were selected and edited using iTOL<sup>25</sup>.

### 138 **Use of GISAID, Microreact, NextStrain, Outbreak.info and Pangolin Resources.**

139 Lineage assignment was performed using the Pangolin online web application<sup>1</sup>. Identification of  
140 additional B.1.1.220 lineage samples with and without the E484K mutation was performed  
141 initially with Microreact<sup>21</sup> and later with GISAID<sup>26</sup>. An acknowledgement of the originating and



142 submitting laboratories providing the SARS-CoV-2 comparator genomes used in this study can  
143 be found in the Supplemental Material (**Table S1** and **S2**).

144 The time-scaled analysis of SARS-CoV-2 genomes (Fig. 1A) was modified from the Nextstrain<sup>27</sup>  
145 SARS-CoV-2 global phylogeny (<https://nextstrain.org/ncov/global>) accessed on March 6<sup>th</sup>, 2021.  
146 This build contained 3,960 SARS-CoV-2 genomes sampled between Dec 2019 and Mar 2021  
147 and collected from: Africa (n = 588), Asia (n = 671), Europe (n = 783), North America (n = 845),  
148 Oceania (n = 489) and South America (n = 594). Full details on bioinformatic processing are  
149 available on the Nexstrain Github page (<https://github.com/nextstrain/ncov>).

150 The USA map of the estimated Lineage B.1.1.220 prevalence (Fig. 1B) was created and edited  
151 from the Outbreak.info<sup>17</sup> dashboard accessed on March 7<sup>th</sup>, 2021.

## 152 **Data submission**

153 The sequences of the five strains described in this paper have been uploaded to GISAID with  
154 Accession ID (virus name): EPI\_ISL\_1184113(hCoV-19/USA/NY-RRHS-101/2021),  
155 EPI\_ISL\_1184114 (hCoV-19/USA/NY-RRHS-102/2021), EPI\_ISL\_1184135 (hCoV-  
156 19/USA/NY-RRHS-114/2021), EPI\_ISL\_1184136 (hCoV-19/USA/NY-RRHS-115/2021), and  
157 EPI\_ISL\_1184137 (hCoV-19/USA/NY-RRHS-120/2021).

158 **Acknowledgements**

159 We are deeply indebted to the clinical and research laboratories of the Rochester Regional Health  
160 System and ACM Global Laboratories, especially Maria Formica, Julie Freed, Chalece Holland,  
161 Jean Campbell, and Julia Nary.

162 This study was partially funded by the Defense Health Program (DHP) Operations &  
163 Maintenance (O&M). Material has been reviewed by the Walter Reed Army Institute of  
164 Research. There is no objection to its presentation. The opinions or assertions contained herein  
165 are the private views of the authors and are not to be construed as official, or reflecting the views  
166 of the Department of the Army or the Department of Defense. Some authors (Bennett, Mc Gann,  
167 Swierczewski) are employees of the U.S. Government and this work was prepared as part of  
168 their official duties.

169 **Bibliography**

- 170 1. Rambaut, A., *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist  
171 genomic epidemiology. *Nat Microbiol* **5**, 1403-1407 (2020).
- 172 2. Houriiyah Tegally., *et al.* Emergence and rapid spread of a new severe acute respiratory  
173 syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in  
174 South Africa. *MedRxiv* (2020). DOI: <https://doi.org/10.1101/2020.12.21.20248640>.
- 175 3. Nuno R., *et al.* Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus,  
176 Brazil. (2021). DOI: <https://doi.org/10.1101/2021.02.26.21252554>.
- 177 4. Rambaut, A., *et al.* Preliminary genomic characterization of an emergent SARS-CoV-2  
178 lineage in the UK defined by a novel set of spike mutations. Unpublished:  
179 [https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-](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)  
180 [lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563](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).
- 181 5. O'Toole, Á *et al.* Pangolin: lineage assignment in an emerging pandemic as an  
182 epidemiological tool. (2021). Unpublished: [github.com/cov-lineages/pangolin](https://github.com/cov-lineages/pangolin)
- 183 6. Davies, N.G., *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage  
184 B.1.1.7 in England. *Science* (2021). Epub ahead of print: 2021/03/05
- 185 7. Volz, S.M., *et al.* Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights  
186 from linking epidemiological and genetic data. *medRxiv* (2021). DOI:  
187 <https://doi.org/10.1101/2020.12.30.20249034>.
- 188 8. Chen, R.E., *et al.* Resistance of SARS-CoV-2 variants to neutralization by monoclonal  
189 and serum-derived polyclonal antibodies. *Nat Med* (2021). Epub ahead of print:  
190 2021/03/06.

- 191 9. Daming Zhou, W.D., *et al.* Evidence of escape of SARS-CoV-2 variant B.1.351 from  
192 natural and vaccine induced sera. *Cell* (2021). DOI:  
193 <https://doi.org/10.1016/j.cell.2021.02.037>.
- 194 10. Wang, W.B., *et al.*. E484K mutation in SARS-CoV-2 RBD enhances binding affinity  
195 with hACE2 but reduces interactions with neutralizing antibodies and nanobodies:  
196 Binding free energy calculation studies. *bioRxiv* (2021). DOI:  
197 <https://doi.org/10.1101/2021.02.17.431566>.
- 198 11. Weisblum, Y., *et al.* Escape from neutralizing antibodies by SARS-CoV-2 spike protein  
199 variants. *Elife* **9** (2020).
- 200 12. Novavax. Novavax COVID-19 Vaccine Demonstrates 89.3% Efficacy in UK Phase 3  
201 Trial. Press release: [https://ir.novavax.com/news-releases/news-release-details/novavax-](https://ir.novavax.com/news-releases/news-release-details/novavax-covid-19-vaccine-demonstrates-893-efficacy-uk-phase-3)  
202 [covid-19-vaccine-demonstrates-893-efficacy-uk-phase-3](https://ir.novavax.com/news-releases/news-release-details/novavax-covid-19-vaccine-demonstrates-893-efficacy-uk-phase-3). (2021).
- 203 13. United States Centers for Disease Control. US COVID-19 Cases Caused by Variants.  
204 <https://www.cdc.gov/coronavirus/2019-ncov/transmission/variant-cases.html>. (2021).
- 205 14. Pater, A.A., *et al.* Emergence and Evolution of a Prevalent New SARS-CoV-2 Variant in  
206 the United States. *bioRxiv* (2021). DOI: <https://doi.org/10.1101/2021.01.11.426287>
- 207 15. Hodcroft, E.B., *et al.* Emergence in late 2020 of multiple lineages of SARS-CoV-2 Spike  
208 protein variants affecting amino acid position 677. *medRxiv* (2021). DOI:  
209 <https://doi.org/10.1101/2021.02.12.21251658>.
- 210 16. Annavajhala, M.K., *et al.* A Novel SARS-CoV-2 Variant of Concern, B.1.526, Identified  
211 in New York. *medRxiv* (2021). DOI: <https://doi.org/10.1101/2021.02.23.21252259>
- 212 17. Gangavarapu, K. *et al.*, Outbreak.info. <https://outbreak.info/> (2020).

- 213 18. Ji, D., *et al.* Prediction for Progression Risk in Patients With COVID-19 Pneumonia: The  
214 CALL Score. *Clin Infect Dis* **71**, 1393-1399 (2020).
- 215 19. Jackson, C.B., Zhang, L., Farzan, M. & Choe, H. Functional importance of the D614G  
216 mutation in the SARS-CoV-2 spike protein. *Biochem Biophys Res Commun* **538**, 108-115  
217 (2021).
- 218 20. Korber, B., *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G  
219 Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827 e819 (2020).
- 220 21. Argimon, S., *et al.* Microreact: visualizing and sharing data for genomic epidemiology  
221 and phylogeography. *Microb Genom* **2**, e000093 (2016).
- 222 22. Kearse M, *et al.*,. Geneious Basic: an integrated and extendable desktop software  
223 platform for the organization and analysis of sequence data. *Bioinformatics*. 2012 Jun  
224 15;28(12):1647-9.
- 225 23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high  
226 throughput. *Nucleic Acids Res*. 2004 Mar 19;32(5):1792-7.
- 227 24. Kozlov AM, *et al.*,. RAxML-NG: a fast, scalable and user-friendly tool for maximum  
228 likelihood phylogenetic inference. *Bioinformatics*. 2019 Nov 1;35(21):4453-4455.
- 229 25. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new  
230 developments. *Nucleic Acids Res*. 2019 Jul 2;47(W1):W256-W259.
- 231 26. Elbe S & Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative  
232 contribution to global health. *Glob Chall*. 2017 Jan 10;1(1):33-46.
- 233 27. Hadfield J, *et al.*,. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*.  
234 2018 Dec 1;34(23):4121-4123.

235

236 **Figure Legend**

237 **Figure 1. Emergence of the E484K mutation in B.1.1.220 SARS-CoV-2 isolates from**

238 **Upstate New York. (A)** Time-scaled, global phylogeny of 3,960 SARS-CoV-2 genomes

239 sampled between Dec 2019 and Mar 2021 (modified from Nextstrain<sup>27</sup> dashboard). A full table  
240 of acknowledgements for the data included in this analysis is available in the supplement (Table

241 S2). Major clades (Nextstrain nomenclature), variants (literature) and lineages (Pangolin) are

242 indicated. Strains are colored based on the presence of a glutamic acid (E, teal) or lysine (K,

243 orange) at position 484 of the spike protein. **(B)** Color-gradient map of the USA showing the

244 estimated percentage of sequenced SARS-CoV-2 genomes that were assigned to Lineage

245 B.1.1.220 since first detection (modified from Outbreak.info<sup>17</sup> dashboard). **(C)** SNP-based,

246 unrooted phylogram of B.1.1.220 variants collected from New York State between October 2020

247 and March 2021 (n = 129). Empty circles indicate strains with Glutamic acid (E) at residue 484

248 of the spike protein while filled circles indicate those with Lysine (K) at that position. Genomes

249 obtained from GISAID<sup>26</sup> (Table S1) are shown in red. Genomes newly described in this report

250 are shown in blue (note that only four out of five circles are visible as two samples from a single

251 patient were identical). **(D)** SNP-based, unrooted maximum-likelihood phylogeny of sixteen

252 monophyletic B.1.1.220 strains carrying the E484K mutation. Both comparator (Table S1) and

253 newly sequenced isolates (bold font) are labeled using their GISAID Accession ID. Where

254 known, the county and collection date are indicated. For both panels C and D, the number of

255 SNPs separating each strain can be inferred from the respective scale.

