



# **Evolutionary and Antigenic Profiling of the Tendentious D614G Mutation of SARS-CoV-2 in Gujarat, India**

Jay Nimavat<sup>1†</sup>, Chandrashekar Mootapally<sup>2†</sup>, Neelam M. Nathani<sup>2†</sup>, Devyani Dave<sup>1</sup>, Mukesh N. Kher<sup>3</sup>, Mayur S. Mahajan<sup>4</sup>, Chaitanya G. Joshi<sup>5</sup> and Vaibhav D. Bhatt<sup>2,4\*</sup>

<sup>1</sup>Department of Pharmaceutical Sciences, Saurashtra University, Rajkot, India, <sup>2</sup>School of Applied Sciences and Technology (GTU-SAST), Gujarat Technological University, Ahmedabad, India, <sup>3</sup>L. M. College of Pharmacy, Ahmedabad, India, <sup>4</sup>Atal Incubation Centre, Gujarat Technological University, Ahmedabad, India, <sup>5</sup>Gujarat Biotechnology Research Centre, Department of Science and Technology, Gandhinagar, India

#### **OPEN ACCESS**

#### Edited by:

Dhaval K. Acharya, B N Patel Institute of Paramedical, India

#### Reviewed by:

Arif Ansori, Airlangga University, Indonesia Alejandro Flores-Alanis, National Autonomous University of Mexico, Mexico

#### \*Correspondence:

Vaibhav D. Bhatt bhatt\_vbhv@yahoo.co.in

<sup>†</sup>These authors have contributed equally to this work

#### Specialty section:

This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 26 August 2021 Accepted: 18 October 2021 Published: 11 November 2021

#### Citation:

Nimavat J, Mootapally C, Nathani NM, Dave D, Kher MN, Mahajan MS, Joshi CG and Bhatt VD (2021) Evolutionary and Antigenic Profiling of the Tendentious D614G Mutation of SARS-CoV-2 in Gujarat, India. Front. Genet. 12:764927. doi: 10.3389/fgene.2021.764927 Humankind has suffered many pandemics in history including measles, SARS, MERS, Ebola, and recently the novel Coronavirus disease caused by SARS-CoV-2. As of September 2021, it has affected over 200 million people and caused over 4 million deaths. India is the second most affected country in the world. Up to this date, more than 38 Lakh viral genomes have been submitted to public repositories like GISAID and NCBI to analyze the virus phylogeny and mutations. Here, we analyzed 2349 genome sequences of SARS-CoV-2 submitted in GISAID by a single institute pertaining to infections from the Gujarat state to know their variants and phylogenetic distributions with a major focus on the spike protein. More than 93% of the genomes had one or more mutations in the spike glycoprotein. The D614G variant in spike protein is reported to have a very high frequency of >95% globally followed by the L452R and P681R, thus getting significant attention. The antigenic propensity of a small peptide of 29 residues from 597 to 625 of the spike protein variants having D614 and G614 showed that G614 has a little higher antigenic propensity. Thus, the D614G is the cause for higher viral antigenicity, however, it has not been reported to be effective to be causing more deaths.

Keywords: antigenic propensity, clades, D614G, SARS-CoV-2, spike glycoprotein

# INTRODUCTION

In the last 2 decades, this is the third instance of a zoonotic coronavirus pandemic. Acute respiratory disease has previously been caused by SARS-CoV, 2002 (Drosten et al., 2003) and MERS-CoV, 2012 (Azhar et al., 2014a; Azhar et al., 2014b) in humans. The novel SARS-CoV-2 virus has recently triggered the coronavirus disease Covid-19. The viral infection began in Wuhan, China in December 2019 and soon became a global outbreak. In a very short span, it has caused significant effects on social and economic activities. Compared to other coronaviruses, this newly emerged SARS-CoV-2 is spreading rapidly, giving challenges to administrative and scientific communities. Influenza, severe respiratory, enteric and neurological complications, elevated white blood cells, and kidney failure are significant indicators of this viral infection. Mammals such as bats are the primary beta coronavirus reservoirs. Due to zoonotic contacts and viral genomic mutations, it is expected to have crossed the species barrier and infected humans. Previous studies indicate that zoonotic infections such as SARS-CoV was transmitted from bats and civets that first infected humans in 2002 (Ksiazek et al., 2003; Marra et al., 2003; Xu et al., 2004). SARS-CoV-2 is enveloped, contains positive

sense ssRNA, and a genome size of 29–30 kb. It belongs to the coronaviridae family and subfamily beta-coronavirus. The family also comprises of MERS CoV which was originated from camel and later led to human transmission in Saudi Arabia (2012) (Azhar et al., 2014a; Chan et al., 2015; Sabir et al., 2016). These infections from bats are predicted to infect humans due to their change in genomic RNA sequence, especially in the spike glycoprotein region (Song et al., 2005; Menachery et al., 2016). Complete genomic sequences of SARS-CoV-2 isolated from infected patients belonging to different geographical locations allows the understanding of these variations and the corresponding influence on viral infecting potency.

SARS CoV-2 similar to SARS-CoV uses angiotensinconverting enzyme II (ACE2) as a receptor for host cell entry. It has spike glycoproteins on the surface, which has two functional domains, S1 and S2. It helps in host cell receptor binding and fusion of viral membrane with the cellular membrane (Harvey et al., 2021). SARS-CoV-2 spike protein has an ACE2 affinity 10 to 20 times greater than that of SARS-CoV spike protein (Walls et al., 2020; Wrapp et al., 2020). Both SARS-CoV and SARS-CoV-2 use CTD (Cterminal domain) of the S1 domain for receptor binding but SARS-CoV-2 binds more strongly than SARS-CoV. Coronaviruses use two different pathways for host cell entry, First, protease mediated cell surface pathway and second, the endosomal pathway. S protein is cleaved into an S1 subunit for receptor binding and an S2 subunit for membrane fusion by the host proteases. Several cellular proteases including furin, transmembrane protease serine 2 (TMPRSS2) and cathepsin (cat) B/L are important for priming SARS-CoV-2 spike protein to enhance ACE2 mediated viral entry (Hoffmann et al., 2020). Spike protein plays a vital role in the evolution of coronaviruses to escape the host immune system. Spike protein shows a higher amount of antigenicity, which is evident from the fact that convalescent plasma from SARS patients shows a high percentage of anti-S neutralizing antibodies (Liu et al., 2006; Wan et al., 2020). There are a lot of variations observed in spike protein sequence, a major variation in spike protein is a non-synonymous D614G mutation (D-Aspartate, G-Glycine) which has received special attention by several groups due to its dominance (Harvey et al., 2021).

Several studies have reported the phylogenomics of the variant and shown that it is leading to higher transmission, though no less influence on the death rate. Few studies have also shown that the variant has increased cellular entry efficacy to human cells compared to the wild type (Wan et al., 2020). In context to the same, here we assessed the antigenic propensity of the epitope encompassing the D614G mutation considering its high frequency and the segment being earlier reported as immune-dominant peptide in SARS-CoV (Wang et al., 2016; Kim et al., 2020). The region-wise analysis of the variant will provide information of this rapidly spreading variant for possible considerations in protective strategy development. Further, we also compiled the current major spike mutations in the Gujarat state in comparison with their global frequencies.

# MATERIALS AND METHOD

# SARS-CoV-2 Sequence Retrieval

A total of 2439 sequences of SARS-CoV-2 genome sequences corresponding to the Gujarat state were retrieved from GISAID (https://www.gisaid.org/) hCov-19 database and these sequences represented different districts of Gujarat state. The genomes were sequenced and submitted by Gujarat Biotechnology Research. The complete genome sequences of SARS-CoV-2 reference genome of the Wuhan isolate (GenBank code: MN908947.3) and Bat CoV-RaTG13 (MN996532.1) were retrieved in FASTA format from NCBI. All the retrieved sequences were subjected to BLAST.

# **Mutation and Phylogenetic Analysis**

Spike protein sequences retrieved from GISAID were aligned with reference spike protein sequences using Jalview version 2.11.1.0 (Waterhouse et al., 2009). Mutations were identified and listed using GISAID EpiCoV<sup>TM</sup> database (Elbe and Buckland-Merrett, 2017). Occurrence of D614G mutation over time was visualized using NextStrain platform (Hadfield et al., 2018) where data is enabled from the GISAID. NextStrain visualization analysis can process up to 3000 genomic sequences at a time. Therefore, for the primary global analysis, they subsample 120 genomes per admin division per month giving results in a more equitable way.

# **Antigenic Propensity Analysis**

A small part of sequence of spike glycoprotein  $S_{597-625}$  from sequences was analyzed using the method described earlier (Kolaskar and Tongaonkar, 1990) provided on an online server by UNIVERSIDAD COMPLUTENSE, MADRID. The interpretation was done as suggested: the average score for the whole protein was used as a cut-off for the then all residues to be considered as potentially antigenic.

# RESULTS

## Sequence Similarity Studies of SARS-CoV-2 Genomes

A total of 2439 sequences of SARS-CoV-2 were retrieved from GISAID platform. Upon performing nucleotide BLAST of reference SARS-CoV-2 with SARS CoV, at the genomic level SARS-CoV-2 and SARS-CoV were observed to have 79.6% sequence similarity. Sequences retrieved from infected patients by SARS-Cov-2 from GISAID had sequence similarity of about 80% with SARS-CoV, 99.9% with SARS-CoV-2 reference sequence, and 96% with the Bat CoV RaTG13.

## Mutation Analysis of SARS-CoV-2 Genomes

Out of 2439 sequences studied, 2400 genomes had a common mutation and there were 9 mutations observed in spike protein that were present in at least 15% of the studied genomes (**Figure 1**). A total of 34 nonsynonymous mutations were observed, with the spike D614G having the highest occurrence in 2400 genomes followed by the nsp12 P4715L mutation observed in 2254 genomes (**Figure 1A**). Nine of the spike mutations had a percentage occurrence in the range of 25–99





FIGURE 2 | (A) Change in the nucleotide at position 23403. NC\_045512.2/1-29903 and MN996532.1/1-29855 are SARS-CoV-2 reference sequence and Bat CoV RaTG13 reference sequence, respectively. (B) Amino acid sequence alignment of spike protein encompassing the D614G position.

in the studied genomes (Figure 1B). After D614G, P681R occurred in 1455 of the analyzed sequences.

All, except D614G, from the 9 non-synonymous mutations observed, had relatively more frequency in the Gujarat region compared to their global frequencies. While D614G had almost similar frequency with its global value. In the genomic sequence for spike glycoprotein, a single mutation, i.e., from A to G nucleotide was prevalent in the majority of genomes at the nucleotide position

number 23403 (Figure 2). This change in the nucleotide causes change in amino acid while translating the gene, because of this aspartate is replaced by glycine in the protein sequence (Figure 2). Its frequency of occurrence has overall increased with the time. Phylogenomics of genomes based on the D and G variants is depicted at the time of initial data collection and the scenario down to the recent timeline (Supplementary Figure S1). Also, there is a clear difference in frequency as observed for D614G



which was comparatively much higher compared to global during early phase up to July 2020 (**Supplementary Figure S2**) and now the frequencies are almost the same. Such difference currently is observed in the P681R mutation, which is the hallmark mutation of the Delta variant wherein the percentage occurrence is 20% higher in the Gujarat region compared to the global. The P681R has outraced other major mutations in spike proteins and is the second major mutation.

We also report here the top 20 nonsynonymous mutations as per global scenario with their frequencies (**Figure 3**). These also reflect that D613G is the most spread, and further there are around 12 mutations that have >25% occurrence globally.

### **Antigenicity of Peptides**

In the present study, we assessed the antigenic propensity of the peptides of the region  $S_{597-625}$  and the results showed that the variant having G at position number 614 is having a little higher antigenic propensity than the one having D at the same position. Isolate from Wuhan (i.e., reference having D614), a Pangolin CoV and a variant having G614 had antigenic propensity of 1.0822, 1.0822, and 1.0824, respectively. In each sequence of 29

residues, a peptide starting from position 10 to the 25th position comprised a single antigenic determinant (**Figure 4**).

## DISCUSSION

A novel corona virus that emerged in December 2019 from Wuhan, China has resulted into a pandemic. In the structure of SARS-CoV-2 virus, spike protein is 1273 amino acid residue in length and forms a trimeric spike on the virion surface. There are many mutations observed in the amino acid composition of the spike protein but primary data shows that strains with S-D614G are more infectious and exhibit high transmission efficiency (Zhang et al., 2020). Regions between amino acid 614 and 621 of SARS-CoV-2 spike proteins were also identified as a B cell epitope by different methods and D614G may affect the antigenicity of this region (Kim et al., 2020). However, there is still much scope to understand how D614G affects antigenic properties of S protein; whether elastase-2 inhibitors and convalescent serum samples of patients can block infection of D614G variant remains unclear.



In the current study, we observed that D614G is highly prevalent mutation in the spike protein of genomes from COVID-19 patients of the Gujarat region. D614 is conserved in the reference sequence from Wuhan and a sequence from Guangzhou, while SARS-CoV-2 genome sequences from Gujarat, India showed a very high frequency of this mutation. In addition, it is concurrently seen with other mutations like P681R, L452R, T19R, E156G, T478K. These had >50% of occurrence in Gujarat whereas at the global level only D614G showed very high occurrence (98.53%) of all sequences and the rest of these formerly mentioned mutations had a frequency of 35–40% globally, lower than that in Gujarat. The observation that the P681R was the second most prevalent mutation reveals the recent high dominance of the Delta variant in the region. Additionally, spike D614G was accompanied by high occurrence of the nsp12 P4715L mutation, and this duo variant which is linked to

pathogenicity was observed to be not linked positively to fatality rates in Africa (Lamptey et al., 2021). Such duo variants need further attention to assess host-based region-specific response.

Considering the high occurrence of D614G in spike protein of SARS-CoV-2, several groups have assessed the antigenic peptides and it is reported that the peptide  $S_{597-625}$  is one of the major immunodominant in humans (Wang et al., 2016; Kim et al., 2020). Antigenic propensity analysis showed that variant spike protein-  $G_{614}$  is having a little higher antigenic propensity to the  $D_{614}$ . This observation may be one of the reasons for no change in the death rate despite the high spread of the variant. Further studies on spike protein epitopes may provide insights on the potential efficacy of many of the vaccines which may be designed based on the D614 sequence.

Earlier reports have also showed that D614G increases the efficiency of cellular entry for the virus across a broad range of

human cell types, including cells from lung, liver, and colon (Daniloski et al., 2021). They also observed that spike- $G_{614}$  is more resistant to proteolytic cleavage during the production of the protein in the host cell.

Seeing the rise in the cases with D614G mutation and its enhanced transmission, the D614G attracts significant consideration by researchers and healthcare field fellows. In the present work, we attempt to report the mutation analysis of spike protein and the antigenic propensity of D614G mutation in the spike protein of the viral isolates from the Gujarat region.

#### DATA AVAILABILITY STATEMENT

The original contributions of SARS-CoV-2 genome sequences (n = 2439) presented in the study are publicly available. This data can be found here: https://www.gisaid.org/, the corresponding accession numbers are available from https://covid.gbrc.res.in/ (Sr. No. 1-2439).

### **AUTHOR CONTRIBUTIONS**

JN: Writing original draft and data analysis. CM: Formal analysis, review, and editing. NN: Formal analysis, review, and editing.

### REFERENCES

- Azhar, E. I., El-Kafrawy, S. A., Farraj, S. A., Hassan, A. M., Al-Saeed, M. S., Hashem, A. M., et al. (2014). Evidence for Camel-To-Human Transmission of MERS Coronavirus. *N. Engl. J. Med.* 370 (26), 2499–2505. PubMed PMID: 24896817. doi:10.1056/NEJMoa1401505
- Azhar, E. I., Hashem, A. M., El-Kafrawy, S. A., Sohrab, S. S., Aburizaiza, A. S., Farraj, S. A., et al. (2014). Detection of the Middle East Respiratory Syndrome Coronavirus Genome in an Air Sample Originating from a Camel Barn Owned by an Infected Patient. *mBio* 5 (4), e01450–14. doi:10.1128/mBio.01450-14
- Chan, J. F. W., Lau, S. K. P., To, K. K. W., Cheng, V. C. C., Woo, P. C. Y., and Yuen, K.-Y. (2015). Middle East Respiratory Syndrome Coronavirus: Another Zoonotic Betacoronavirus Causing SARS-like Disease. *Clin. Microbiol. Rev.* 28 (2), 465–522. doi:10.1128/CMR.00102-14
- Daniloski, Z., Jordan, T. X., Ilmain, J. K., Guo, X., Bhabha, G., tenOever, B. R., et al. (2021). The Spike D614G Mutation Increases SARS-CoV-2 Infection of Multiple Human Cell Types. *Elife* 10, e65365, 2021. PubMed PMID: 33570490. doi:10.7554/eLife.65365
- Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., et al. (2003). Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. N. Engl. J. Med. 348 (20), 1967–1976. PubMed PMID: 12690091. doi:10.1056/NEJMoa030747
- Elbe, S., and Buckland-Merrett, G. (2017). Data, Disease and Diplomacy: GISAID's Innovative Contribution to Global Health. *Glob. Challenges* 1 (1), 33–46. doi:10.1002/gch2.1018
- Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., et al. (2018). Nextstrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics* 34 (23), 4121–4123. doi:10.1093/bioinformatics/bty407
- Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., et al. (2021). SARS-CoV-2 Variants, Spike Mutations and Immune Escape. *Nat. Rev. Microbiol.* 19 (7), 409–424. doi:10.1038/s41579-021-00573-0
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181 (2), 271–280e8. doi:10.1016/j.cell.2020.02.052

DD: Data collection. MK: Data collection. MM: Manuscript formatting and data analysisv. CJ: Conceptualization, writing—review and editingv. VB: Conceptualization, writing—review and editingv. All authors contributed to the manuscript and approved the submitted version.

#### ACKNOWLEDGMENTS

The authors are thankful to Gujarat Biotechnology Research Centre for providing the data of Covid-19 samples.

### SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.764927/full#supplementary-material

Supplementary Figure S1 | Phylogenomic distribution of D and G variants (A) upto the studied genomes viz., July 2020 and (B) scenario down to the recent timeline viz., September 2021.

**Supplementary Figure S2** | Non synonymous mutations observed in the spike glycoprotein sequence of analysed genomes from Gujarat (n = 315) compared to the globally submitted genome sequences of SARS-CoV-2 compared to Wuhan reference by EpiCoVTM on GISAID platform as on June 30, 2020.

- Kim, S.-J., Nguyen, V.-G., Park, Y.-H., Park, B.-K., and Chung, H.-C. (2020). A Novel Synonymous Mutation of SARS-CoV-2: Is This Possible to Affect Their Antigenicity and Immunogenicity? *Vaccines* 8 (2), 220, 2020. PubMed PMID:. doi:10.3390/vaccines8020220
- Kolaskar, A. S., and Tongaonkar, P. C. (1990). A Semi-empirical Method for Prediction of Antigenic Determinants on Protein Antigens. *FEBS Lett.* 276 (1-2), 172–174. doi:10.1016/0014-5793(90)80535-Q
- Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., et al. (2003). A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. N. Engl. J. Med. 348 (20), 1953–1966. PubMed PMID: 12690092. doi:10.1056/NEJMoa030781
- Lamptey, J., Oyelami, F. O., Owusu, M., Nkrumah, B., Idowu, P. O., Adu-Gyamfi, E. A., et al. (2021). Genomic and Epidemiological Characteristics of SARS-CoV-2 in Africa. *Plos Negl. Trop. Dis.* 15 (4), e0009335. doi:10.1371/ journal.pntd.0009335
- Liu, W., Fontanet, A., Zhang, P. H., Zhan, L., Xin, Z. T., Baril, L., et al. (2006). Two-Year Prospective Study of the Humoral Immune Response of Patients with Severe Acute Respiratory Syndrome. *J. Infect. Dis.* 193 (6), 792–795. doi:10.1086/500469
- Marra, M. A., Jones, S. J. M., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S. N., et al. (2003). The Genome Sequence of the SARS-Associated Coronavirus. *Science* 300 (5624), 1399–1404. doi:10.1126/science.1085953
- Menachery, V. D., Yount, B. L., Sims, A. C., Debbink, K., Agnihothram, S. S., Gralinski, L. E., et al. (2016). SARS-like WIV1-CoV Poised for Human Emergence. *Proc. Natl. Acad. Sci. USA* 113 (11), 3048–3053. doi:10.1073/pnas.1517719113
- Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., et al. (2003). Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *Science* 300 (5624), 1394–1399. doi:10.1126/ science.1085952
- Sabir, J. S., Lam, T. T., Ahmed, M. M., Li, L., Shen, Y., Abo-Aba, S. E., et al. (2016). Cocirculation of Three Camel Coronavirus Species and Recombination of MERS-CoVs in Saudi Arabia. *Science* 351 (6268), 81–84. doi:10.1126/science.aac8608
- Song, H.-D., Tu, C.-C., Zhang, G.-W., Wang, S.-Y., Zheng, K., Lei, L.-C., et al. (2005). Cross-host Evolution of Severe Acute Respiratory Syndrome Coronavirus in palm Civet and Human. *Proc. Natl. Acad. Sci.* 102 (7), 2430–2435. doi:10.1073/pnas.0409608102

- Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181 (2), 281–292e6. doi:10.1016/j.cell.2020.02.058
- Wan, Y., Shang, J., Graham, R., Baric, R. S., and Li, F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* 94 (7), e00127–20. doi:10.1128/jvi.00127-20
- Wang, Q., Zhang, L., Kuwahara, K., Li, L., Liu, Z., Li, T., et al. (2016). Immunodominant SARS Coronavirus Epitopes in Humans Elicited Both Enhancing and Neutralizing Effects on Infection in Non-human Primates. ACS Infect. Dis. 2 (5), 361–376. doi:10.1021/acsinfecdis.6b00006
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2--a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* 25 (9), 1189–1191. doi:10.1093/bioinformatics/ btp033
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *Science* 367 (6483), 1260–1263. doi:10.1126/ science.abb2507
- Xu, R.-H., He, J.-F., Evans, M. R., Peng, G.-W., Field, H. E., Yu, D.-W., et al. (2004). Epidemiologic Clues to SARS Origin in China. *Emerg. Infect. Dis.* 10 (6), 1030–1037. doi:10.3201/eid1006.030852

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., et al. (2020). SARS-CoV-2 Spike-Protein D614G Mutation Increases Virion Spike Density and Infectivity. *Nat. Commun.* 11 (1), 6013–6019. doi:10.1038/s41467-020-19808-4

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Nimavat, Mootapally, Nathani, Dave, Kher, Mahajan, Joshi and Bhatt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.