AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

Rapid SARS-CoV-2 Adaptation to Available Cellular Proteases

M. Zeeshan Chaudhry,[a]* Kathrin Eschke,[a] Markus Hoffmann,[b,c] Martina Grashoff,[d] Leila Abassi,[a] Yeonsu Kim,[a] Linda Brunotte,[e] Stephan Ludwig,[e] Andrea Kröger,[d,f] Frank Klawonn,[g,h] Stefan H. Pöhlmann,[b,c] Luka Cicin-Sain[a,i]

[a]Department of Viral Immunology, Helmholtz Centre for Infection Research, Braunschweig, Germany
[b]Infection Biology Unit, German Primate Center, Göttingen, Germany
[c]Faculty of Biology and Psychology, Georg-August-University Göttingen, Göttingen, Germany
[d]Research Group Innate Immunity and Infection, Helmholtz Centre for Infection Research, Braunschweig, Germany
[e]Institut für Virologie (IMV), Westfälische Wilhelms-Universität Münster, Münster, Germany
[f]Institute of Medical Microbiology and Hospital Hygiene, Otto von Guericke University, Magdeburg, Germany
[g]Biostatistics Group, Helmholtz Centre for Infection Research, Braunschweig, Germany
[h]Department of Computer Science, Ostfalia University, Wolfenbüttel, Germany
[i]Centre for Individualized Infection Medicine (CIIM), Hannover, Germany

M. Zeeshan Chaudhry and Kathrin Eschke contributed equally to this article. Author order was determined by the corresponding author after negotiation.

ABSTRACT Recent emergence of SARS-CoV-1 variants demonstrates the potential of this virus for targeted evolution, despite its overall genomic stability. Here we show the dynamics and the mechanisms behind the rapid adaptation of SARS-CoV-2 to growth in Vero E6 cells. The selective advantage for growth in Vero E6 cells is due to increased cleavage efficiency by cathepsins at the mutated S1/S2 site. S1/S2 site also constitutes a heparan sulfate (HS) binding motif that influenced virus growth in Vero E6 cells, but HS antagonist did not inhibit virus adaptation in these cells. The entry of Vero E6-adapted virus into human cells is defective because the mutated spike variants are poorly processed by furin or TMPRSS2. Minor subpopulation that lack the furin cleavage motif in the spike protein rapidly become dominant upon passaging through Vero E6 cells, but wild type sequences are maintained at low percentage in the virus swarm and mediate a rapid reverse adaptation if the virus is passaged again on TMPRSS2$^+$ human cells. Our data show that the spike protein of SARS-CoV-2 can rapidly adapt itself to available proteases and argue for deep sequence surveillance to identify the emergence of novel variants.

IMPORTANCE Recently emerging SARS-CoV-2 variants B.1.1.7 (alpha variant), B.1.617.2 (delta variant), and B.1.1.529 (omicron variant) harbor spike mutations and have been linked to increased virus pathogenesis. The emergence of these novel variants highlights coronavirus adaptation and evolution potential, despite the stable consensus genotype of clinical isolates. We show that subdominant variants maintained in the virus population enable the virus to rapidly adapt to selection pressure. Although these adaptations lead to genotype change, the change is not absolute and genomes with original genotype are maintained in the virus swarm. Thus, our results imply that the relative stability of SARS-CoV-2 in numerous independent clinical isolates belies its potential for rapid adaptation to new conditions.

KEYWORDS SARS-CoV-2, furin cleavage site, coronavirus spike priming, spike mutation, deep sequencing

Coronavirus disease COVID-19 emerged in the city of Wuhan in December 2019 and since has killed more than 3.5 million people worldwide. The pandemic spread of the disease that affected the human society was found to be caused by a novel betacoronavirus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is closely related to severe acute respiratory syndrome coronavirus (SARS-CoV-1) (1, 2). The typical, crown-like appearance of the coronaviruses (CoVs) is attributed to homotrimers of spike (S) protein (3),

Editor Tom Gallagher, Loyola University Chicago

Address correspondence to M. Zeeshan Chaudhry, uqmzees1@uq.edu.au, or Luka Cicin-Sain, luka.cicin-sain@helmholtz-hzi.de.

*Present address: M. Zeeshan Chaudhry, University of Queensland Diamantina Institute, University of Queensland, Woollongabba, Queensland, Australia.

The authors declare no conflict of interest.

Received 23 December 2021
Accepted 2 January 2022
Accepted manuscript posted online 12 January 2022
Published 9 March 2022

which mediate virion binding to the host cell receptors and subsequent membrane fusion. The CoV S protein can be subdivided into two functional subunits: S1 and S2 (4). In the case of SARS-CoV-2, the S1 subunit directly binds to angiotensin-converting enzyme 2 (ACE2), which has been identified as the cellular receptor for SARS-CoV-1 and SARS-CoV-2 (5, 6). Upon ACE2 binding, S needs to be cleaved by cellular proteases at the S1/S2 junction and the S2′ site (5, 7). This exposes the fusion peptide in the S2 subunit, allowing the fusion of viral and cellular membranes.

CoVs can enter the cell by fusing the virus envelope with cytoplasmic membrane at either the cell surface or the endosome, depending on the localization and availability of cellular proteases in the host cell. Hence, cellular proteases have a direct impact on the cellular tropism and pathogenesis of CoVs. Different CoVs have evolved multiple ways to achieve proteolytic priming. A wide diversity of cellular proteases, including trypsin, endosomal cathepsins, transmembrane serine proteases (e.g., TMPRSS2), and furin are known to be involved in the priming of CoV spike (7). The SARS-CoV-2 S protein harbors a multibasic cleavage site at S1/S2, as opposed to the monobasic site that is present in the SARS_CoV-1 S (5). The multibasic site at S1/S2 also constitutes a putative furin cleavage site (RRAR). The consensus sequence for furin cleavage site (FCS) is widely described as R-X-K/R-R (8, 9). The S protein of numerous betacoronaviruses can be activated by furin at S1/S2 site; here, furin typically recognizes and cleaves at the RRXRR motif (7, 10). SARS-CoV-1 and murine hepatitis virus strain 2 (MHV-2) are two prominent exceptions of the betacoronavirus family that lack a FCS and need to be cleaved by other cellular proteases (7).

Interestingly, the furin recognition sequence present in SARS-CoV-2 S is identical (however, in opposite orientation) to the putative heparan sulfate (HS) interaction consensus sequences (XBBXBX, where B is a basic amino acid) (11). It has been shown that cell culture-adapted Sindbis virus can mediate attachment to cell surface HS via furin cleavage site (8). Similarly, upon *in vitro* culturing, CoVs often exhibit a trade-off between the furin cleavability of S protein and HS binding (10, 12). Thus, some CoVs can adapt to use HS as entry receptor in cultured cells (12). Recently, SARS-CoV-2 mutants with deletions at the S1/S2 junction have been described to emerge upon culturing the virus in Vero E6 cells (13). Clinical SARS-CoV-2 isolates, identified during the course of the current pandemic, have shown remarkably few mutations in the consensus sequence at the furin cleavage (14). However, several newly emerged SARS-CoV-2 variants seem to harbor mutations in the receptor binding domain (RBD) of the S protein and exhibit increased transmission. For instance, the SARS-CoV-2 variant B.1.1.7 that emerged in the United Kingdom, or the B1.617 variant from India, have been associated with surges of COVID-19 cases (15, 16).

In order to understand the dynamics of SARS-CoV-2 adaptation, we serially passaged virus strains in defined cell types and controlled environment. We show that SARS-CoV-2 rapidly adapts to culture conditions by natural selection of pre-existing subdominant populations in the virus swarm. Furthermore, we show that the loss of the furin cleavage motif in Vero cells is due to a selection for virus variants with S protein optimized for efficient cleavage with cathepsins. However, viruses with the intact motif are retained as subdominant population in the swarm. Therefore, reverting the passaging of such mutated swarms onto cells expressing TMPRSS2 results in a prompt reversion to the original genotype observed in low-passage clinical isolates.

RESULTS

SARS-CoV-2 rapidly adapted upon culture on Vero E6 cells. We performed deep sequencing of minimally passaged SARS-CoV-2 genomes isolated from virus particles to investigate the genome diversity and presence of minor variants. We observed considerable diversity in Braunschweig (Br) strain upon *in vitro* passage (Br-P2), where a total of 13,211 genomic positions showed mutation frequency that was higher than 0.1%, (Fig. 1A). Similarly, NK-P4 showed around 12,000 nucleotide positions (12,406) with minor variants at $>$0.1% frequency. In order to exclude that the observed diversity is due to *in vitro* passaging, we also analyzed publicly available SARS-CoV-2 RNA-Seq data (GSA accession number CRA002390) from bronchoalveolar lavage (BAL) samples (17). SARS-CoV-2 sequences from two patients showed high

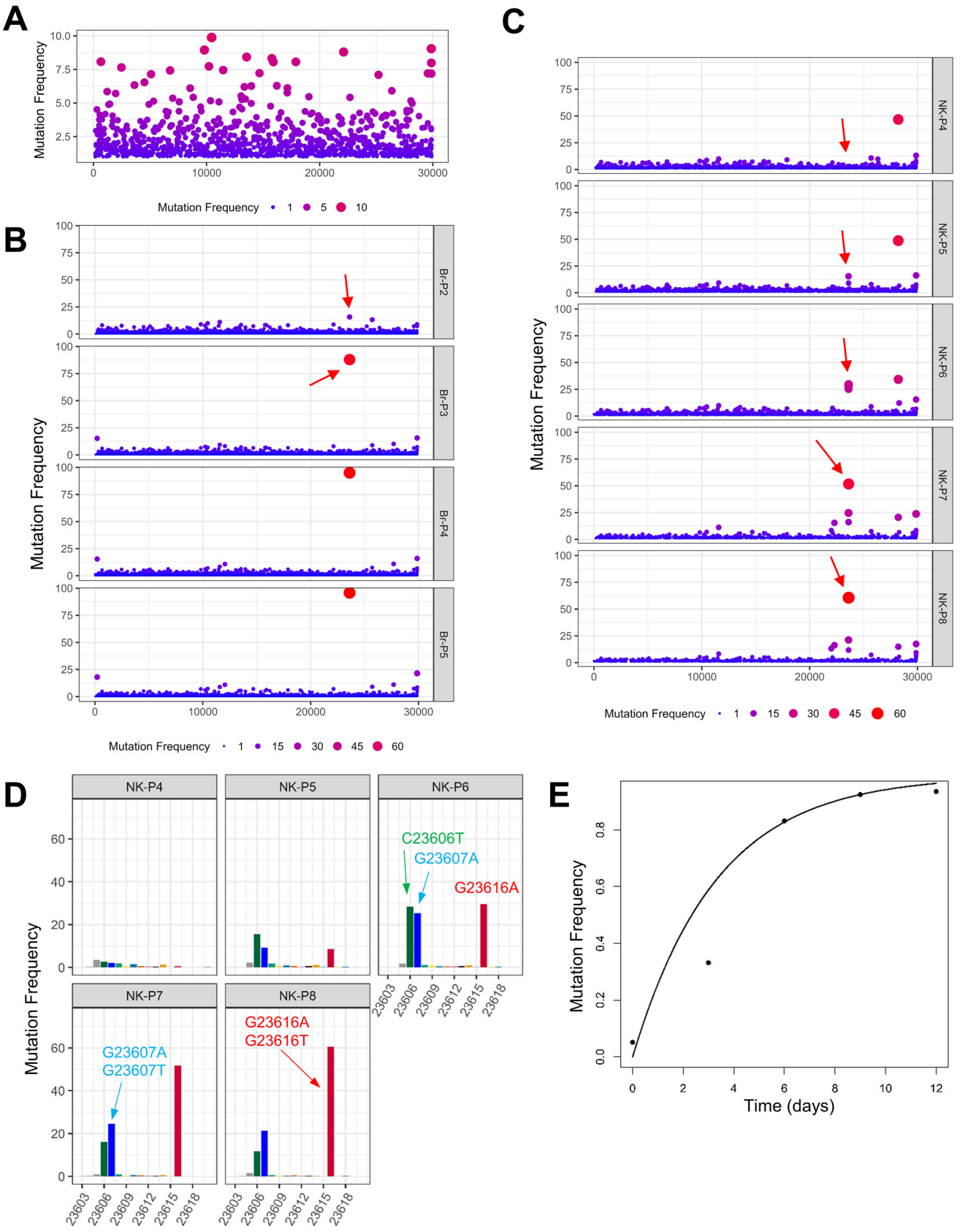

FIG 1 SARS-CoV-2 rapidly adapts upon passage in cultured cells. (A) SARS-CoV-2 with minimal *in vitro* passaging (Br-P2) was analyzed with deep sequencing and genome diversity is highlighted by plotting genome positions with >1% mutation frequency. (B, C) Serial passages of Braunschweig (Br) or

(Continued on next page)

diversity akin to our minimally passaged isolates. Patient 1 sequence had 17,410 positions with >0.1% mutation frequency, while SARS-CoV-2 genome from patient 2 had 6,920 positions with >0.1% allele frequency. On the other hand, a bacterial artificial chromosome (BAC) derived SARS-CoV-2 clone has been shown to possess relatively low diversity and lower number of genome position with >0.1% mutation frequency (18). These data suggested that SARS-CoV-2 maintains quasispecies diversity in natural conditions despite a stable consensus genotype and that our sequencing data are not a result of *in vitro* artifacts.

Sequence analysis of different SARS-CoV-2 clinical isolates passaged on Vero E6 cells revealed point mutations that resulted in the loss of the furin cleavage site PRRARS at the S1/S2 junction region (Table S1), in line with other publications (13, 19). This indicated that mutations of the PRRARS motif were advantageous to the mutated viruses. To understand how fast such mutations occur, we serially passaged SARS-CoV-2 clinical isolates on Vero E6 cells and sequenced their genomes. Overall, the SARS-CoV-2 genomes were stable after multiple passages. However, we observed a very rapid onset of mutations in the furin recognition motif at the S1/S2 site upon passaging in Vero E6 cells (Fig. 1B to D). The NK strain developed multiple point mutations at the furin cleavage site (C23606T, G23607A, and G23616A), resulting in a loss of arginine at positions 682 and/or 685 (Fig. 1C and D). The NK strain passage 6 showed C23606T, G23607A, and G23616A as dominant point mutations at FCS, but other mutations, such as G23607T and G23616T, emerged later in passage 8 virus population. The NK strain also showed an emergence of mutations at low frequency in other parts of the S protein sequence; these were for instance the N148K (11.59%) and H245R (16.26%) mutations observed after eight passages in Vero E6 cells (Fig. 1C). The mutation frequency for N148K and H245R increased at passage 12 to 64.11% and 19.25%, respectively, in the NK-P12 virus stock (Fig. S2), yet these mutations were not observed in other strains and hence were not followed further. We used the Br isolate to compare the dynamics of the S1/S2 site mutation because we had access to earlier passages of this strain. A swift change in the composition of viral populations was observed in this isolate, where >85% of passage 3 (P3) genomes showed a point mutation (C23606A) in the S protein at R682 (Table S1; Fig. S1A). We used another isolate (FI) to compare the nature of mutations at the furin cleavage site and observed a single point mutation (C23606T) in the RRAR site (Fig. S1C), which lead to amino acid change identical to the one in the Br strain. Rapid mutations in the furin site were accompanied with the retention of a small fraction of genomes with wild type furin motif in all strains (Table S1). For instance, the NK strain retained ~8.35% genomes with the intact wild type furin site after 12 passages in Vero E6 (Fig. S2).

A mathematical model was used to fit a non-linear curve to observed mutation frequency data and calculate the rate of mutation frequency change at the furin cleavage site (Fig. 1E). The NK strain showed presence of multiple sub-species including three major variants with nucleotide mutations at genome position 23,606, 23,607, and 23,616 (Fig. S1D). Thus, we summed frequencies of mutations at these three positions to model the change in mutation frequency. For NK strain, the rate was calculated as 0.3 (change in mutation frequency/day). The Br strain mutated at a much higher rate of 0.81 per day (Fig. S1B). We attribute this to the high mutation frequency present in the Br strain at the earliest passage (Br-P2).

The rapid adaption of independent isolates and mutations at the furin cleavage site could have been due to *de novo* mutations, or by the presence of minor variants in the original clinical isolate. However, the depth of sequencing in the initial isolate is necessarily constrained due to the availability of sample material before *in vitro* virus expansion. This introduces a bias that impairs direct comparisons of genomic diversity between original isolates and passaged virus. To overcome this bias, and address the contribution of *de novo* mutations to virus diversification, we plaque purified an isolate from low-passaged

FIG 1 Legend (Continued)
Ischgl (NK) strain were analyzed to assess the composition of viral populations upon passaging in Vero E6 cells. Each symbol represents an individual nucleotide, and genomic positions (*x* axis) with mutation frequency >1% are plotted. Red arrows highlight the position of the furin cleavage site in the genome. (D) The mutation frequency of nucleotides at the furin cleavage site in the NK strain at indicated passage in Vero E6 cells are plotted on the *y* axis. Each bar represents one nucleotide for corresponding genomic position from 23,603 to 23,620 (*x* axis). Dominant mutations are highlighted for certain positions. (E) The sum of mutation frequencies at the furin site position 23,606, 23,607, and 23,616 of NK strain serially passaged on Vero E6 cells is plotted against time. The genomic positions correspond to Wuhan-Hu-1 isolate (GenBank accession no: NC_045512).

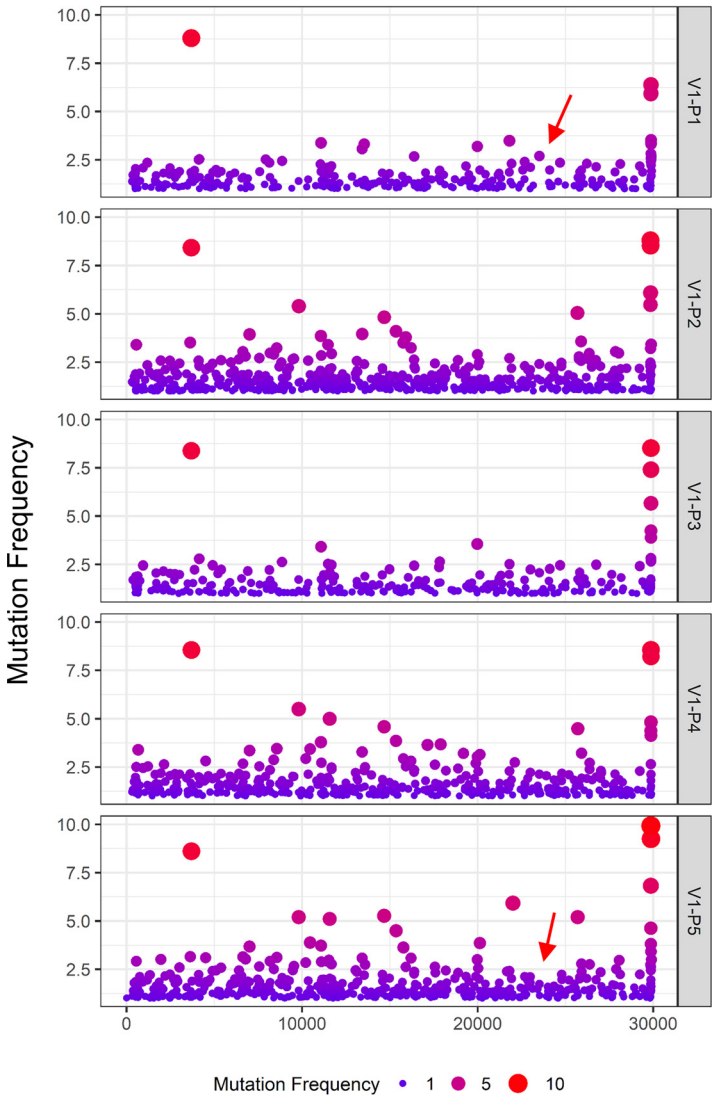

FIG 2 Rapid adaption of SARS-CoV-2 is hampered due to absence of subdominant variants at FCS. Plaque purified virus of FI strain was passaged on Vero E6 and virus genomes were analyzed sequentially with deep sequencing to observe genome diversity at each passage. Red arrows highlight the position of the furin cleavage site in the genome. The genomic positions on *x* axis correspond to Wuhan-Hu-1 isolate (GenBank accession no: NC_045512).

FI-P4 strain, ascertained that the population has the furin cleavage site intact and analyzed the dynamics of mutations. Virus populations derived from a single virion showed considerable genomic diversity (Fig. 2; Fig. S3), but importantly, not at the furin cleavage site. At this position, two out of three tested virus populations showed a completely mutated furin cleavage site, whereas the third clone was uniformly maintaining the wild type allele (at least, to the extent of our sequencing depth). We then passaged the virus strain with the wild type furin site and observed no detectable mutations at the S1/S2 site within five passages on Vero E6 cells. These findings are in line with published data (18), and further corroborated the assumption that sub-dominant variants in virus population enable rapid virus adaption to selection pressure. Overall, the data demonstrated that the S1/S2 furin cleavage site was rapidly mutated upon SARS-CoV-2 culture in Vero E6 cells, but it was not completely lost and virus genomes with the wild type furin site were retained as subdominant variants in the population.

High-passage SARS-CoV-2 strains showed improved growth on Vero E6 but impaired growth on Calu-3 cells. In order to understand if the observed mutations in S may influence viral growth, we compared the cell-to-cell virus spread on Vero E6 cells by assessing plaque

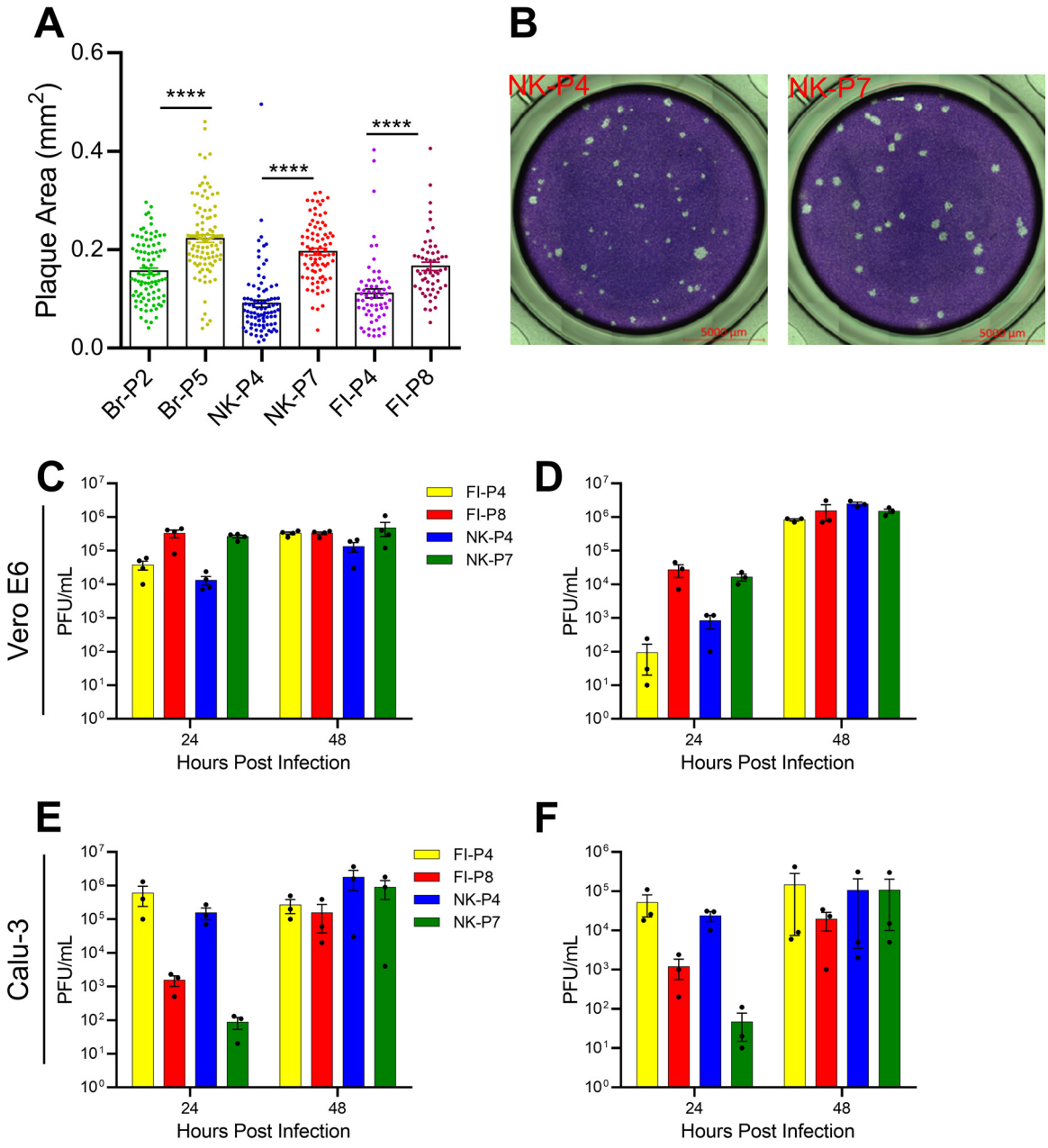

FIG 3 Growth properties of passaged SARS-CoV-2 strains in different cell lines. (A, B) Vero E6 cells were infected with indicated passages of SARS-CoV-2 strains and the plaque size was measured at 72 hpi. Panel A shows the area of virus plaques on Vero E6 cells, and panel B shows representative images of two different wells infected with NK-P4 and NK-P7. (C, D) Virus growth kinetics on Vero E6 cells were established by infecting the cells at an MOI of 0.001. Supernatants (panel C) and cell lysates (panel D) were collected at indicated time points and titrated on Vero E6 cells. (E, F) Calu-3 cells were infected at an MOI of 0.001. Supernatants (panel E) and cell lysates (panel F) were collected at indicated time postinfection and titrated on Vero E6 cells. Data are representative of two independent experiments. Means ±SEM are plotted. Each symbol in panel A represent one plaque and data are pooled from multiple infected wells of two independent experiments. Panel C-F symbols represent biological replicates. Statistical significance was calculated using one way ANOVA with Bonferroni post-test. $^{**}P < 0.01$; $^{****}P < 0.0001$.

sizes. At 72 hours post infection (hpi), a significant increase in plaque sizes of high-passage virus was observed in three independently passaged virus strains (Fig. 3A and B). To validate the results independently, we tested multi-step virus growth kinetics of low- or high-passage isolates NK and FI in Vero E6 cells and noticed that in both cases the high-passage virus showed significantly higher titers at 24 hpi (Fig. 3C). We considered the possibility that the high-passage virus could be more efficiently released from infected cells, and thus compared virus titers in infected cells lysates (Fig. 3D). In both conditions, the high-passage viruses

showed higher titers early upon infection, but this difference in titers was abolished by 48 hpi. Taken together, we concluded that the growth advantage of high-passage virus was a result of an early event that occurred before virus release.

To ascertain the growth properties of Vero-passaged virus on cells expressing TMPRSS2, we tested the growth of virus strains on Calu-3 cells, a cell line derived from human respiratory epithelium, and observed a mirror image of growth properties observed in Vero E6 cells. The low-passage viruses grew substantially better at early time points postinfection and the effects were seen equally in cell lysates and in the supernatants (Fig. 3E and F). These results agree with the previous report that a multibasic cleavage site in the S protein is necessary for activation by TMPRSS2 (20).

SARS-CoV-2 rapidly restored the furin cleavage site upon passage on TMPRSS2$^+$ cells. Low-passage virus grew faster in Calu-3 cells than high-passaged one, but the high-passaged virus caught up with the low-passaged one by 48 hpi (Fig. 3E and F). This suggested that the furin cleavage site confers growth advantage in these cells, but also that it may drive rapid selection of SARS-CoV-2 genomes with the wild type furin site. Therefore, the NK strain that had been passaged six times in Vero E6 cells (NK6) was repassaged on Calu-3 cells and sequenced to analyze S1/S2 site mutations. SARS-CoV-2 genomes with wild type furin site were present as a subdominant fraction ($\sim$15%) in the NK6 viral population. Upon four passages in Calu-3 cells, we observed a substantial reversion to the wild type sequence with the intact furin site ($>$97%) (Fig. 4A). A similar reversion was observed upon passaging the virus on Caco-2 cells, a human gut epithelium cell line expressing TMPRSS2. The sequence changed to a functional furin cleavage site in the vast majority ($>$95%) of viral genomes (Fig. 4A). The data suggested that viruses that demonstrably lost the furin cleavage site rapidly changed the consensus genotype back to wild type in human TMPRSS2$^+$ cells. Thus, to understand the kinetics of the reversion, the virus was sequenced after each sequential passage on Calu-3. Interestingly, the majority of the reversion occurred already by the first passage (Fig. 4B), but a fraction of genomes with the mutated cleavage site was retained as subdominant population in the virus swarm. The reversion rate (change in mutation frequency/day) of the NK6 virus population was 0.63 (Fig. 4C). The low-passage NK virus (NK4) consisted of a population that retained the furin motif in most of the variants, and passaging on Calu-3 or Caco-2 cells did not alter the population considerably (Fig. 3D; Fig. S4A). The rapid initial rate of mutations in both directions suggested that the furin site mutated as a result of targeted adaptation, rather than drift. Furthermore, all viruses passaged on Calu-3 or Caco-2 cells showed a highly similar and stable population composition, regardless of the underlying mutations induced by prior passaging on Vero cells. Interestingly, hitherto unreported mutations L27S and T30I in the envelope (E) protein occurred repeatedly upon virus passaging in Calu-3 and in Caco-2 cells (Fig. S4B, C; Table S2), but this line of research was not pursued further.

Role of the SARS-CoV-2-S furin cleavage site in cell entry. Cellular proteases are critical for SARS-CoV-2 cell entry and their availability affects the virus tropism (5). Therefore, to test the role of different proteases in virus replication, we infected Vero E6, Calu-3, and Caco-2 cells with low- and high-passage viruses in the presence of a TMPRSS2 inhibitor (camostat), a furin inhibitor (F-I), or of a broad-spectrum inhibitor of endosomal proteases (E-64d). The growth of all virus strains was inhibited by E-64d on Vero E6 cells, but not by camostat or F-I, alone or in combination (Fig. 5A). The combined use of E-64d and other protease inhibitors did not depress virus titers over the values observed in cells treated with E-64d alone (Fig. 4A and Fig. S4A). This suggests that endosomal proteases, but not the cell-surface ones, were crucial for SARS-CoV-2 replication in Vero E6 cells. Importantly, no difference in titer between the high- and the low-passage virus was observed in the presence of E-64d, arguing that the dominant mutations present in the high-passaged strains affected the virus entry process in Vero E6 cells. The endosomal protease inhibitor abrogated the growth advantage of mutated viruses in Vero E6 cells and both low- and high-passaged viruses showed similar levels of growth in the presence of inhibitor.

An inverse phenotype was observed in Calu-3 cells, where the TMPRSS2 inhibitor reduced the growth of all tested viruses, whereas E-64d and F-I showed a less pronounced effect (Fig. 5B). The passage-dependent differences in virus titers, seen in untreated controls,

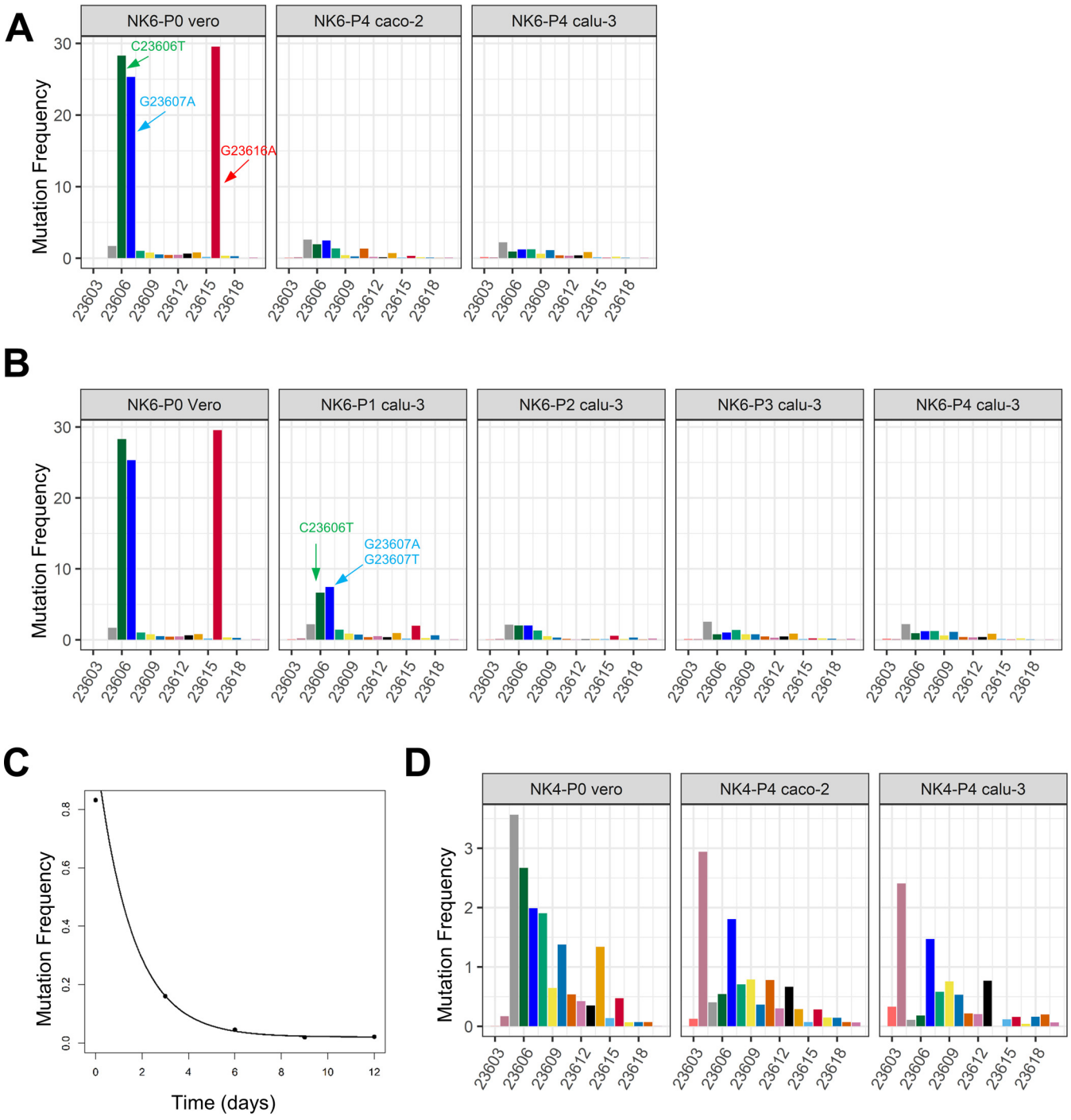

FIG 4 Selection of SARS-CoV-2 subdominant variants with furin cleavage site in TMPRSS2$^+$ human cells. The mutation frequency of genomic positions 23,603 to 23,620 (*x* axis) in SARS-CoV-2 genome corresponding to furin cleavage site are plotted in panels A, B, and D. Dominant mutation in mutated population are highlighted with arrows. (A) NK strain passaged 6 time in Vero E6 cells (NK6) was serially passaged independently on Calu-3 and Caco-2 cells. P0 vero represents the genome of input virus and P4 calu-3 and P4 caco-2 are the viruses cultured on respective cell lines for four passages. (B) The change in SARS-CoV-2 furin site mutations is shown by plotting the nucleotide mutation frequency of NK6 virus upon serial passaging in Calu-3 cells. (C) The sum of mutation frequency at furin site position 23,606, 23,607, and 23,616 of NK6 strain serially passaged on Calu-3 cells is plotted against time and fitted to the mathematical model. (D) NK strain passaged four times in Vero E6 cells (NK4) was serially passaged on Calu-3 and Caco-2 cells. P0 vero represents the genome of low-passage input virus and P4 calu-3 and P4 caco-2 are the derivative viruses serially passaged on Calu-3 and Caco-2 cell lines for four additional cycles.

were erased in the presence of camostat. Finally, virus growth in the presence of all three inhibitors was entirely abrogated in Calu-3 cells (Fig. S5A). Virus replication in Caco-2 cells was not affected by inhibition of any protease alone, but combinations of protease inhibitors involving camostat resulted in a reduction of virus titers (Fig. 5C). Although at 48 hpi the

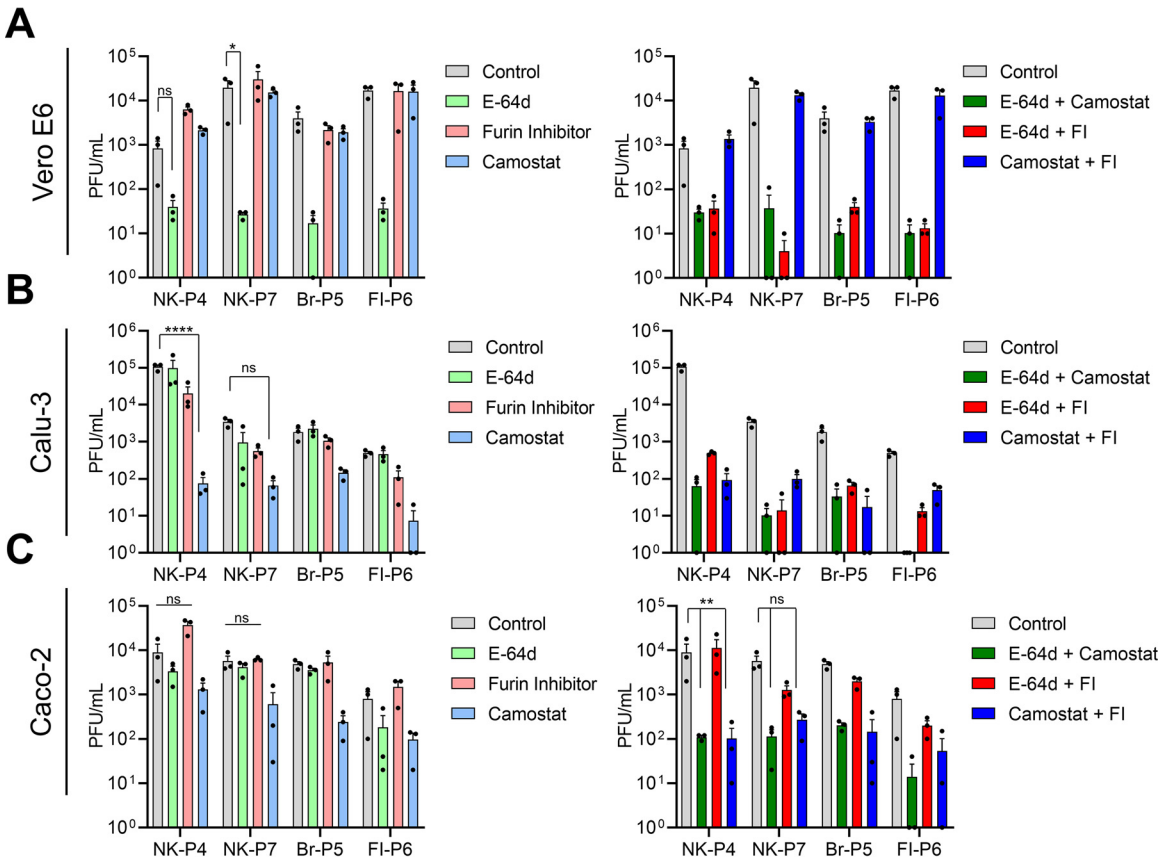

FIG 5 Growth inhibition of low and high-passage SARS-CoV-2 by protease inhibitors. Vero E6 (A), Calu-3 (B), and Caco-2 (C) cells were infected at an MOI of 0.01 in the presence of indicated protease inhibitors. The supernatant from infected cells was collected at 24 hpi and titrated on Vero E6 cells. All experiments were twice performed independently with similar results and a typical data set is shown. Three biological replicates per condition are shown as individual dots. Histogram show means and error bars represent ± SEM. Statistical significance was calculated using two-way ANOVA with Dunnett posttest, where untreated control cells served as reference. ns – $P > 0.05$, *$P < 0.05$, **$P < 0.01$, ****$P > 0.0001$.

growth of low- and high-passage viruses was similar in untreated controls, the pattern of virus growth reduction in the presence of protease inhibitors was similar to the one seen at 24 hpi (Fig. S4B, D).

Heparan sulfate antagonist did not inhibit SARS-CoV-2 adaptation in Vero E6 cells. It has been previously reported that CoVs may trade furin cleavage sites in cell culture for a HS binding site, thus improving virus binding to the cell surface (10, 12). However, in the case of SARS-CoV-2, the furin site PRRARS has a composition that overlaps with the HS binding motif XBBXBX. The observed mutations in HS binding motif (R682W and R685H) decreased the positive charge of the site and increased its hydrophobicity. We hypothesized that stronger PRRARS binding to HS would lead to virus retention on cell surface and hinder cell entry. The observed mutations in HS binding motif may decrease the binding of the S protein to the HS on the plasma membrane, thus enhancing the virus entry in TMPRSS2- Vero E6 and, consequently, the growth in this cell type. To test this hypothesis, we compared the growth of the low-passage virus with the high-passage mutants with HS antagonists surfen or lactoferrin in Vero E6 cells. At 24 hpi, we observed a significant 10-fold median increase in virus titers of the low-passage virus in the presence of surfen. This effect was nonsignificant for Vero E6 adapted viruses (Fig. 6A). Furthermore, absorbing the virus on soluble or immobilized heparin resulted in a modest decrease in virus infectiousness of the low-passaged virus, but no effects on the entry of high-passages virus (Fig. S6A). On the other hand, the virus plaque size was significantly increased upon HS antagonist treatment in both the low- and in the high-passaged viruses (Fig. 6B; Fig. S6B), arguing against a role of HS binding in virus adaptation. Thus, to conclusively clarify this phenomenon, we passaged the viruses in the presence of surfen and sequenced the genomes after four passages. We observed that the S1/S2 site was mutated in

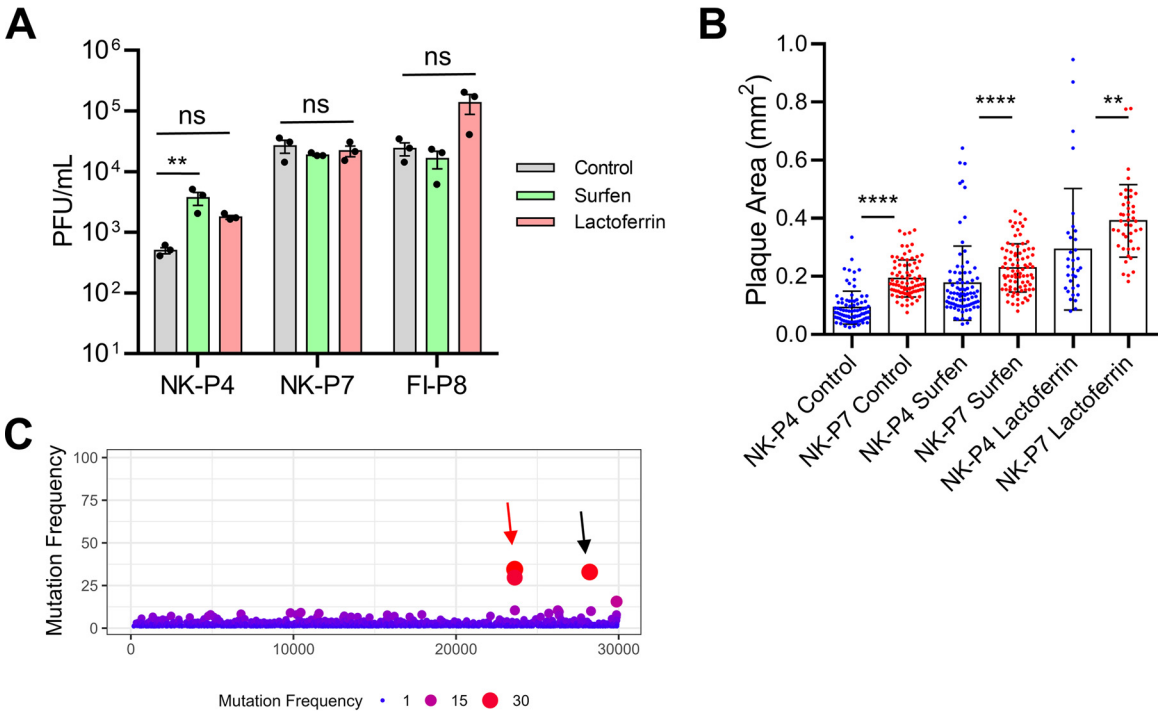

FIG 6 Heparan sulfate antagonists fails to inhibit SARS-CoV-2 adaptation on Vero E6 cells. (A) Vero E6 cells were infected with indicated SARS-CoV-2 passaged strains at an MOI of 0.001 and treated with surfen or lactoferrin. The supernatant was collected at 24 hpi and titrated on Vero E6 cells. Data are representative of two independent experiments, and each symbol represents a biological replicate. Mean and ±SEM is plotted. (B) Vero E6 cells were infected with low- or high-passage NK strain viruses. The cells were overlaid with methylcellulose supplemented with 10 μM surfen or 1 mg/mL lactoferrin. The virus plaque size was quantified 72 hpi. Each symbol represents one plaque and data is pooled from multiple infected wells of two independent experiments. (C) NK strain was passaged in Vero E6 cells in the presence of 10 μM surfen. After four passages, virus genome sequence was analyzed with deep sequencing. Each symbol represents an individual nucleotide, and genomic positions (*x* axis) with mutation frequency >1% are plotted. Red arrow highlights the position of the furin cleavage site and the black arrow shows a dominant silent mutation. Statistical significance was calculated using one way ANOVA test and Bonferroni post-test. ns – $P > 0.05$; **$P < 0.01$; ****$P > 0.0001$.

three independent strains passaged in the presence of the HS antagonist (Fig. 6C; Fig. S6C, D). Therefore, our data indicates that although HS antagonists modestly influence the growth of low-passaged SARS-CoV-2 virus, heparan sulfate binding does not drive the SARS-CoV-2 adaptation in Vero E6 cells.

SARS-CoV-2 S1/S2 site adaption to available proteases. We next investigated if the observed mutations at the S1/S2 site changed the cleavage efficiency of specific proteases. We used peptides derived from the S1/S2 site of SARS-CoV-2 and labeled them fluorescently with FAM-TAMRA FRET pairs to assess the peptide cleavage efficiency by different proteases (Fig. 7A). We compared mutated peptides with P**W**RARS or PRRA**H**S sequences, the two major mutated S variants that were detected in the passaged strains, with the wild type peptide at the S1/S2 site PRRARS. The wild type peptide was rapidly cleaved by furin, but the mutated peptides were not (Fig. 7B). The peptide cleavage kinetics showed that the P**W**RARS peptide site was processed at a high rate by both recombinant cathepsins used in our assays, while the PRRA**H**S variant was more efficiently cleaved by cathepsin B than the wild type peptide. Interestingly, trypsin processed the PWRARS peptide with higher efficiency than the wild type variant (Fig. 7C). We concluded that the motif at the S1/S2 site from high-passaged SARS-CoV-2 was more efficiently cleaved by cathepsins, arguing for their importance in virus cell entry into Vero E6 cells. Thus, efficient cathepsin cleavage is a highly plausible explanation of the selection of SARS-CoV-2 S variants passaged on Vero E6 cells.

Cell entry of SARS-CoV-2-S harboring pseudoviruses. Given that the peptides of S variants with mutated S1/S2 were more efficiently processed by cathepsins, we assumed that TMPRSS2 would be less efficient at processing the mutated spike proteins. Thus, we generated VSV pseudotyped viruses that harbored different SARS-CoV-2 S variants. We analyzed the VSV pseudovirus entry in Vero E6, Calu-3, and Caco-2 cells to ascertain the role of furin

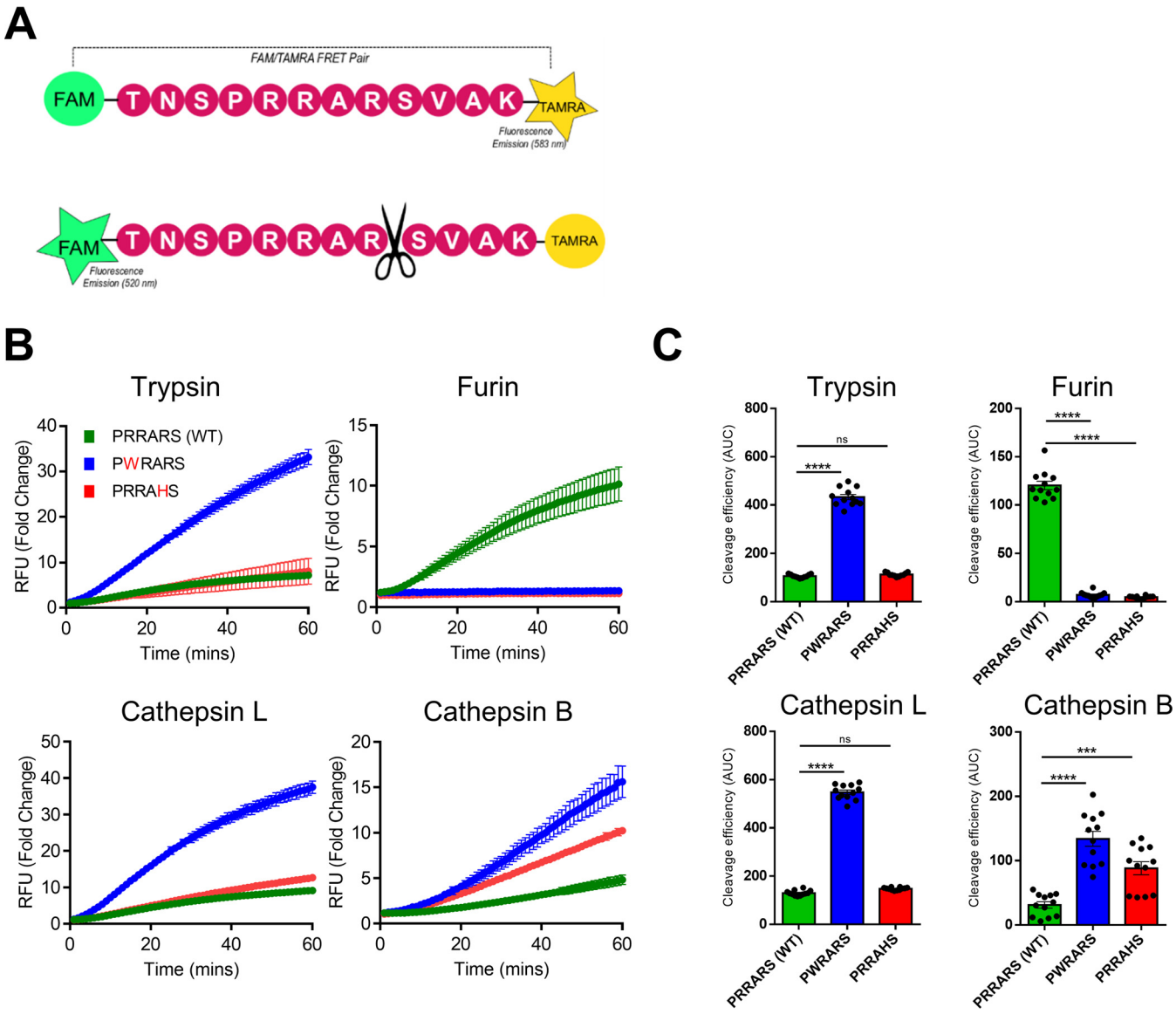

FIG 7 SARS-CoV-2 mutated spike variants are more efficiently cleaved by cathepsins. (A) Design of the FAM-TAMRA fret pair coupled by the mimetic peptide representing the S1/S2 spike cleavage site. (B) Fluorogenic S1/S2 spike cleavage peptides were cleaved by recombinant proteases *in vitro* and fold-increase in FAM florescence over time is shown. Data is representative of three independent experiments, and means ± SD of four replicates are plotted. (C) Peptide cleavage efficiency by different recombinant proteases was measured by calculating the area under the curve (AUC) from panel B curves. Technical replicates from three different experiment are pooled and shown as individual dots. Histograms and error bars are means ± SEM. Statistical significance was calculated with one way ANOVA and Bonferroni post-test. ns, $P > 0.05$; ***$P < 0.001$; ****$P > 0.0001$.

cleavage site in cell entry processes. Viruses harboring mutated SARS-CoV-2 S showed enhanced entry in Vero E6 cells, but the entry of these variants in Calu-3 and Caco-2 was substantially less efficient than the entry of wild type SARS-CoV-2 S virus (Fig. 8A), which was consistent with sensitivity to TMPRSS2 cleavage (Fig. 5B and C). We next tested if S variants were proteolitically cleaved upon infection. An HA tag was attached to the S2 domain of the spike variants in pseudoviruses, which were grown on furin expressing HEK cells and tested by HA-specific Western blot (WB). We noticed that only the wild type S variant showed a shift in the WB, indicative of proteolytic cleavage (Fig. 8B), whereas the mutated variants showed the band at a size corresponding to the full-length S. We concluded that the growth enhancement of passaged SARS-CoV-2 in Vero E6 cells is due to the mutations at the S1/S2 site that modulates the virus entry in these cells.

Mutated S1/S2 site is not cleaved by furin and hampers efficient cell-cell fusion. The PRRARS motif present at S1/S2 site of SARS-CoV-2 S matches the furin consensus

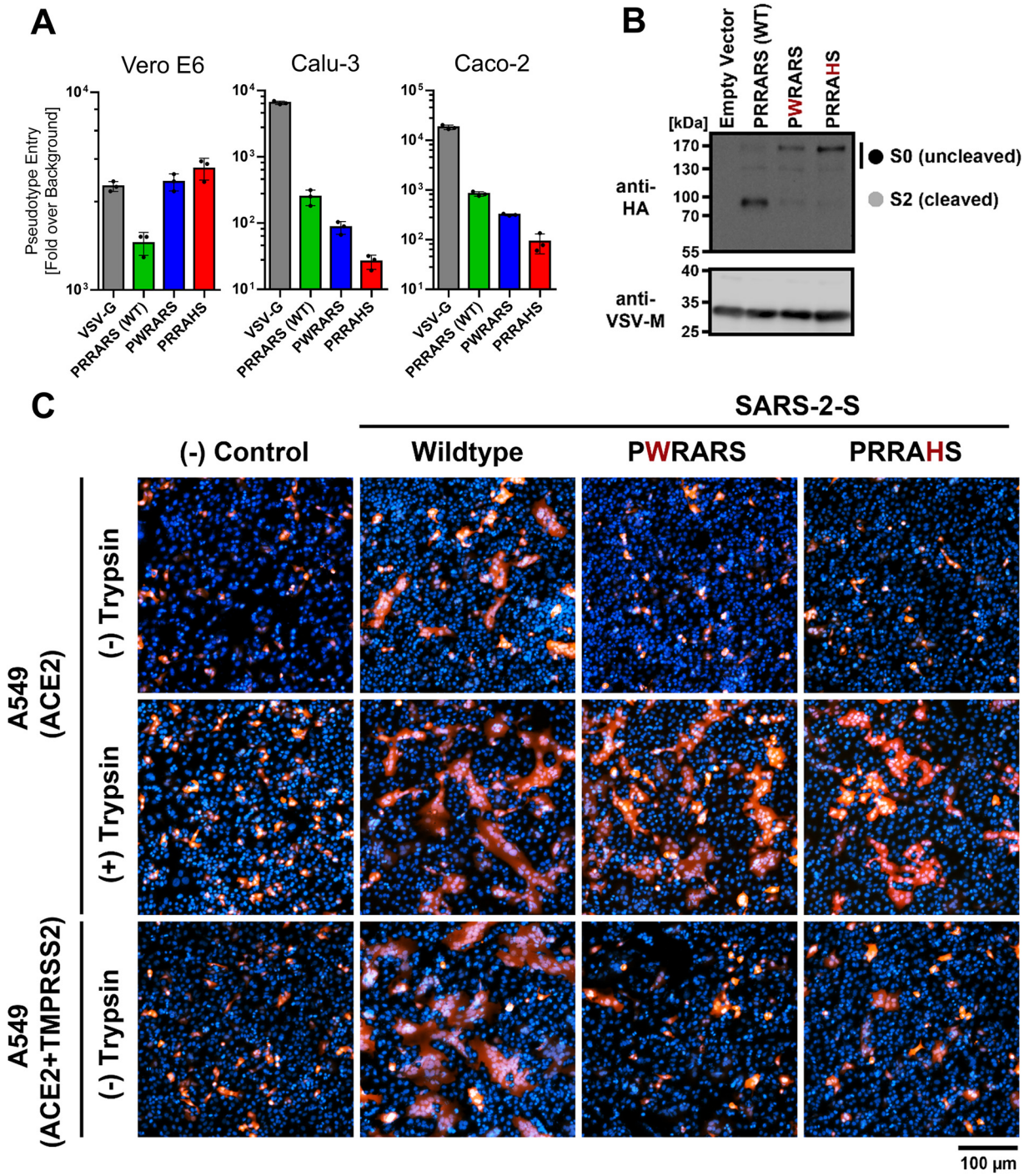

FIG 8 SARS-CoV-2 S furin cleavage site is required for syncytium formation and cell entry into TMPRSS2^{+} human cells. (A) Vero E6, Calu-3, and Caco-2 cells were infected with pseudotyped VSV harboring VSV-G, wild type, or mutated SARS-CoV-2 spike protein. At 16 hpi, pseudotype entry was analyzed by determining luciferase activity. Signals obtained for particles bearing no envelope protein were used for normalization. The average of three independent experiments is plotted along with ± SEM. (B) Analysis of furin-mediated S protein cleavage. Rhabdoviral particles harboring the indicated S proteins containing a C-terminal HA-tag for detection were lysed and subjected to Western blot analysis. Detection of vesicular stomatitis virus matrix protein (VSV-M) served as control. (C) Syncytium formation assay: A549-ACE2 or A549-ACE2-TMPRSS2 cells were co-transfected with DsRed expressing plasmid and vector that expressed the indicated S proteins (or no S protein, (–) Control). At 24 hpi, cells were incubated in the presence or absence of trypsin (1 μg/mL) for 1 h, before they were fixed, stained with DAPI and analyzed by florescent microscopy (scale bars, 100 μm).

sequence RX[K/R]R and can be efficiently cleaved by furin (20), but mutated S proteins with PWRARS or PRRAHS sequence were not (Fig. 8B). We next investigated if the mutated S can mediate syncytia formation in A549 cells, a cell type that expresses furin, albeit at low levels (21). We cotransfected cells with plasmids that expressed DsRed and S protein and observed that wild type S protein expression resulted in multi-nucleated giant cell formation (Fig. 8C), likely due to the furin expression in A549 cells. Mutated S proteins induced a few tiny syncytia in the absence of trypsin treatment. However, trypsin treatment increased syncytia size in all conditions. Lastly, we showed that TMPRSS2 expression enhanced cell-cell fusion and syncytia size by wild type S protein, but the effect was clearly less pronounced in the case of mutated S proteins.

DISCUSSION

RNA virus populations are composed of swarms of different genome variants known as quasispecies (22, 23) that arise due to erroneous proof-reading of RNA-dependent RNA polymerase (RdRp), and are essential for adaptive evolution and fitness of RNA viruses (24). Our deep sequencing data showed that SARS-CoV-2 exhibits remarkable genome stability overall. However, this stability does not hinder the ability of the virus to adapt quickly. Furthermore, high diversity in the clinical SARS-CoV-2 isolates compared with BAC-derived clones suggests that these minor variants play an important role in overall virus fitness and its adaptation potential (18).

Although the loss of furin cleavage site led to virus growth advantage in Vero E6 cells, genomes with the wild type PRRARS motif were not completely eliminated. The virus population accumulated the mutations at the furin cleavage site in an asymptotic manner, stabilizing after four to five passages and maintaining a subdominant fraction of the population with the intact motif. Mutation dynamics in Calu-3 cells followed a similar, albeit reverse pattern, and viruses with mutated PRRARS motifs were retained in the virus swarm. We observed different adaptation rates in different strains and the velocity of adaptation positively correlated with the size of the subdominant population in these strains, which may indicate that adaptation dynamics are likely dependent on the presence and size of subdominant populations in the virus swarm. The lack of rapid emergence of S1/S2 mutations in BAC-derived SARS-CoV-2 clones (18) and virus populations derived from single virion demonstrate that the rapid emergence of mutations that we observed is likely due to the presence of minor variants in clinical isolates. These observations suggest that SARS-CoV-2 maintains a considerable diversity in the quasispecies, facilitating natural selection and a rapid virus adaptation to changing environment and conditions. The imperfect fitness at the individual level adds to the ability of the population to quickly adapt, thus contributing to the overall fitness of the swarm.

Our results are in agreement with the observation that the presence of a multibasic site is necessary for SARS-CoV-2 spike processing by furin and TMPRSS2 (20). Once the arginine at position 685 is replaced by histidine, which is another basic amino acid, to form the PRRAHS mutated site, the ratio of hydrophilic residues change from 67% to 50%. Hence, although PRRAHS fulfills multibasic site criteria, the arginine at position 685 is required for TMPRSS2 spike cleavage. It is unclear if the PRRAHS site is cleaved by cathepsin-B at the histidine or at one of the other arginine residues.

The growth advantage on Vero E6 cells was surprising, because the furin cleavage site should have provided an advantage to viruses on cells that express furin or TMPRSS2 enzymes (Calu-3 or Caco-2), but its absence should not have conferred growth advantage in cells that lack these enzymes. SARS-CoV-2 furin cleavage site PRRARS has a composition that overlaps with the HS binding motif XBBXBX, and HS binding in the absence of cleavage enzymes might exert selection pressure. However, the importance of these HS binding receptors *in vivo* is debated as many viruses acquire HS binding upon cell culture adaptation (25). The SARS-CoV-2 S protein has been shown to bind HS (26) and heparin was shown to inhibit SARS-CoV-2 infection *in vitro* (27). Furthermore, Kim et al. has identified PRRARS site in the SARS-CoV-2 S protein as a putative HS binding site using an unbiased ligand-docking model (26). We report that the putative HS binding motif influences the growth of SARS-CoV-2, but virus adaptation is not affected by HS antagonists. Instead, the

virus adaptation is driven by selection of S1/S2 variants that can be more efficiently cleaved by cathepsins, which are responsible for spike cleavage and virus entry in Vero E6 cells. The reverse adaptation to original genotype in Caco-2 or Calu-3 cells is likely due to TMPRSS2 dependent cleavage because they differ in levels of cathepsin expression, but show a similar outcome (28).

Our study suggests that clinical SARS-CoV-2 isolates should be deep sequenced to identify minor genome variants present in the virus swarm, rather than solely focusing on the consensus genotype. The emergence of novel SARS-CoV-2 variants with enhanced entry in human lung cells highlights that this virus can rapidly adapt (29). Furthermore, variants with reduced neutralization by sera from convalescent patients or vaccinated individuals were observed in populations with high SARS-CoV-2 seroprevalence (30). Our results argue that these variants could have been present as minor subpopulations in individual hosts and expanded because of selection pressure. Overall, we demonstrate a high potential of SARS-CoV-2 rapid adaptation, due to its swarm-like replication. Future research in animal models should explore the potential for the *in vivo* adaptation of SARS-CoV-2, the quasispecies diversity, and the bottlenecks imposed on the virus spread. Our research indicates that the presence of subdominant genetic variants within SARS-CoV-2 isolates needs to be considered as a potential determinant of their virulence.

MATERIALS AND METHODS

Cell cultures and viruses. Vero E6 (ATCC CRL-1586) and 293T (DSMZ ACC-635) cells were maintained in Dulbecco's modified eagle medium (DMEM) medium supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 IU/mL penicillin, and 100 μg/mL streptomycin. Vero-TMPRSS2 cells stably expressing human TMPRSS2 were generated previously by retroviral transduction and blasticidin-based selection (5). Calu-3 (ATCC HTB-55) and Caco-2 (ATCC HTB-37) were cultured in Eagle's minimum essential medium (EMEM) supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 IU/mL penicillin, 100 μg/mL streptomycin, and 1x non-essential amino acid solution (Gibco MEM Non-Essential Amino Acids Solution 100X). A549 (ATCC CCL-185) cells were incubated in DMEM/F-12 medium with Nutrient Mix (ThermoFisher Scientific). All incubations of cells and virus were at 37°C in a 5% CO_2 atmosphere.

The SARS-CoV-2 strains used in the study are Braunschweig isolate (hCoV-19/Germany/Br-ZK-1/2020, GISAID database ID: EPI_ISL_491115), South Tyrol isolate (hCoV-19/Germany/Muenster_FI1103201/2020, GISAID database ID: EPI_ISL_463008), and Ischgl isolate (hCoV-19/Germany/NK1103201/2020).

SARS-CoV-2 passage and virus stock generation. Braunschweig isolate (Br) was derived from an oropharyngeal swab using Vero E6 cells. Ischgl (NK) and South Tyrol (FI) strains were isolated by Stephan Ludwig lab in Muenster. All strains were passaged twice in Vero E6 cells at an MOI (multiplicity of infection) of 0.01 to obtain working virus stocks. For serial passaging of SARS-CoV-2 strains, Vero E6, Calu-3, and Caco-2 cells were seeded in T25 or T75 flasks 1 day before infection. Confluent monolayers of the cells were infected with virus and virus supernatant was collected 3 hpi.

Initial low-passage virus stocks were generated in a two-step protocol, where a seeding stock was generated by infecting one T75 flask of Vero E6 cells. The seeding stock virus supernatant was collected 3 hpi and used to infect 10 T75 flasks for virus stock generation. The virus stock supernatant was collected from all flasks at 3 hpi and spun at 3,000 g for 10 min to remove cell debris. Then, the virus supernatant was concentrated using Vivaspin 20 concentrators (Sartorius Stedim Biotech) by spinning at 6,000 g for 30 min. The resulting virus stock was aliquoted and stored at −80°C until further use.

SARS-CoV-2 titration and plaque purification. SARS-CoV-2 was titrated with virus plaque assay on Vero E6 cells. Virus stocks were serially diluted in virus titration media (VTM, DMEM supplemented with 5% fetal calf serum, 2 mM L-glutamine, 100 IU/mL penicillin, and 100 μg/mL streptomycin) and titrated on 24-well plates. The virus inoculum was added on the Vero E6 cells and incubated at 37°C. After 1 h, the inoculum was removed and the cell were overlaid with VTM supplemented with 1.5% carboxymethylcellulose (medium viscosity, C9481, Sigma-Aldrich) and incubated at 37°C. The overlay was removed from the cell at 3 hpi and the plates were fixed by submerging them in a tank of 6% formaldehyde (methanol stabilized) for at least 1 h. The cell monolayer was stained with crystal violet and the plaques were quantified by visual inspection with microscope. Virus supernatants were titrated in a similar fashion on 96-well plates.

For plaque purification, Vero E6 cells seeded in 24-well plate were infected with 10 PFU of early passage FI-P4 strain. After 1 h of infection, the inoculum was removed, and cells were overlaid with VTM supplemented with 1.5% carboxymethylcellulose and incubated at 37°C for 3 days. Plaques were identified under the inverted microscope and well-separated plaques were picked by scratching with 200-μL pipette tips. The tips were transferred to Vero E6 cells in a T25 flask to recover the virus.

Growth kinetics and plaque area measurement. Vero E6, Calu-3, or Caco-2 were seeded in a well of a 96-well plate. Confluent monolayers were infected with at an MOI of 0.001. After 1 h of infection at 37°C, the inoculum was removed and replaced by normal media. Supernatants and infected cell lysates were collected at different time points postinfection and virus titer were determined as described above.

To analyze cell-to-cell spread of SARS-CoV-2, the virus plaque area was determined. Vero E6 seeded in a well of a 24-well plate were infected with 30 PFU of SARS CoV 2. After 1 h of infection, the inoculum was removed and cells were overlaid with VTM supplemented with 1.5% carboxymethylcellulose and

incubated at 37°C. For testing the effect of HS antagonist, cells were pretreated with 10 μM surfen or 1 mg/mL lactoferrin for 1 h at 37°C before infection, and infection was performed in the presence of surfen or lactoferrin. Finally, after 1 h, the virus inoculum was removed and monolayer was overlaid with VTM supplemented with 1.5% carboxymethylcellulose and 10 μM surfen or 1 mg/mL lactoferrin. Cells were fixed and stained as described above and plaque areas were quantified using a Zeiss LSM 980 Microscope. The plaque area was measured with help of ZEISS ZEN lite 3.0 (Blue edition) and plotted using GraphPad Prism.

SARS-CoV-2 infection in the presence of protease inhibitors and HS antagonists. SARS-CoV-2 cell entry inhibition was tested in 96-well plates, by pretreating cells with protease inhibitors (E-64d, 10 μM; camostat, 10 μM; furin inhibitor, 10 μM) or HS antagonists (10 μM surfen or 1 mg/mL lactoferrin) for 1 h at 37°C before virus infection. The infection was performed in the presence of protease inhibitors or HS antagonists, and after 1 h infection at 37°C, the virus inoculum was removed and normal media supplemented with inhibitors or HS antagonists was added to the cells. The virus supernatant was collected at 1 and 2 hpi and titrated on Vero E6 cells as described in the SARS-CoV-2 titration section.

For analyzing the heparin binding, ~1,000 PFU of SARS-CoV-2 were incubated with heparin sodium salt (100 μg/mL) for 1 h at 37°C. After incubation, the virus was titrated on Vero E6 cells in a 24-well plate format. Similarly, for testing the ability of immobilized heparin to bind SARS-CoV-2, Heparin−biotin sodium salt was immobilized on Pierce Streptavidin Coated Plates (Thermo Fisher Scientific). The volume used to immobilize Heparin-biotin (100 μg/mL) was 100 μL per well. Plates were incubated at 37°C for 15 min and then washed three times with 1x PBS. The virus suspension (1,000 PFU) was added to the heparin-coated wells along with empty streptavidin coated wells and incubated for 1 h at 37°C. Afterwards, the suspension was titrated on Vero E6 cells in a 24-well plate format.

Pseudovirus entry assay. Vesicular stomatitis virus (VSV) pseudotyped with SARS-CoV-2 S were generated according to a published protocol (31). In brief, 293T cells were transfected with expression plasmids for SARS-CoV-2 S proteins of either wild type Wuhan/Hu-1/2019 (lineage B, with D614G mutation inserted) or variants with mutations in the furin cleavage site. At 24 h posttransfection, the transfection medium was removed and cells were inoculated with a replication-deficient VSV vector lacking the genetic information for the VSV glycoprotein (VSV-G), and instead coding for an enhanced green fluorescent protein and firefly luciferase from independent transcription units, VSV*ΔG-FLuc (kindly provided by Gert Zimmer, Institute of Virology and Immunology, Mittelhäusern, Switzerland) (32). Following 1 h of incubation at 37°C and 5% CO2, the inoculum was removed and cells were washed with PBS, before culture medium containing anti-VSV-G antibody (culture supernatant from I1-hybridoma cells; ATCC CRL-2700) was added and cells were further incubated. The VSV pseudotype virus particles (VSVpp) were harvested at 16 to 18 hpi. The culture medium was collected, centrifuged (2,000 g, 10 min, RT) to pellet cellular debris and the clarified supernatant was transferred into fresh tubes and stored at −80°C until further use. Each batch of pseudotypes was pretested for comparable transduction efficiencies by the respective S proteins and absence of transduction by bald control particles before being used. Furthermore, for each construct an untagged variant as well as a version containing a C345 terminal HA epitope tag was constructed.

Western blot analysis. For the analysis of S protein processing, we subjected VSVpp harboring HA-tagged S proteins to SDS-PAGE and Western blot analysis. For this, we loaded 1 mL VSVpp onto 50 μL of a 20% (wt/vol) sucrose cushion and performed high-speed centrifugation (25.000 g, 120 min, 4°C). Next, we removed 1 mL of supernatant, added 50 μL of 2 $\times$ SDS-sample buffer and incubated the samples for 15 min at 96°C. Thereafter, the samples were subjected to SDS-PAGE and protein transfer to nitrocellulose membranes by Western blotting. The membranes were subsequently blocked in 5% skim milk solution (PBS containing 0.05% Tween 20 [PBS-T] and 5% skim milk powder) for 1 h at room temperature. The blots were then incubated over night at 4°C with primary antibody solution (all antibodies were diluted in PBS-T containing 5% skim milk and mouse anti-HA tag [Sigma-Aldrich, H3663, 1:2,500] or VSV matrix protein [Kerafast, EB0011, 1:2,500]). Following this incubation, the blots were washed 3x with PBS-T before they were incubated for 1 h at room temperature with peroxidase-coupled goat anti-mouse antibody (Dianova, 115-035-003, 1:10,000). Finally, the blots were again washed and imaged. For this, an in-house-prepared enhanced chemiluminescent solution (0.1 M Tris-HCl [pH 8.6], 250 μg/mL luminol, 1 mg/mL para-hydroxycoumaric acid, 0.3% H2O2) and the ChemoCam imaging system along with the ChemoStar Professional software (Intas Science Imaging Instruments GmbH) were used.

Syncytium formation assay. A549-ACE2 or A549-ACE2-TMPRSS2 cells were grown on coverslips seeded in 24-well plates and cotransfected with DsRed vector (1 μg/well) and S protein expression plasmid (1 μg/well) using Lipofectamine 2000 LTX with Plus reagent (Thermo Fisher Scientific) and Opti-MEM medium (Gibco). After 6 h, the transfection solutions were removed and the cells further incubated for 24 h in standard culture medium. The medium was changed next day to serum-free medium with or without 1 μg/mL bovine trypsin (Sigma-Aldrich) and the cells were incubated for additional 1 h at 37°C. Cells were washed with PBS and fixed with 4% paraformaldehyde solution for 20 min at room temperature. After fixation, cells were stained with DAPI solution and analyzed by bright-field microscopy using a Zeiss LSM800 confocal laser scanning microscope and the ZEN imaging software.

Peptide cleavage assay. Target cleavage peptides derived from the SARS-CoV-2 spike S1/S2 site amino acid sequence TNSPRRARSVA and flanked with FAM-TAMRA FRET pair ([FAM]TNSPRRARSVA-K[TAMRA] [COOH.acetate]) were synthesized by Thermo Fisher Scientific. Wild type peptide TNSPRRARSVA and two mutated peptides TNSPWRARSVA and TNSPRRAHSVA were synthesized. Recombinant furin was purchased from New England Biolabs. Recombinant L-1-Tosylamide-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin was obtained from Sigma-Aldrich. Recombinant cathepsin B and cathepsin L were purchased from R&D Systems.

Florescent peptides were diluted to 25 μM and the cleavage reaction was performed in 100 μL total volume. The reactions buffer was composed of 100 mM HEPES, 0.5% Triton X-100, 1 mM CaCl2, and

1 mM 2-mercaptoethanol (pH 7.5) for furin (10 U/mL); PBS for trypsin (100 ng/mL); 25 mM MES, pH 5.0 for cathepsin B (100 ng/μL); and 50 mM MES, 5 mM DTT, 1 mM EDTA, 0.005% (wt/vol) Brij-35, pH 6.0 for cathepsin L (100 ng/μL). Reactions were performed in Varioskan Flash (Thermo Scientific) at 30°C, and kinetic fluorescence measurements were recorded at a 1-min interval for 60 min with λex 495-nm and λem 520-nm wavelengths setting to measure FAM emissions after cleavage of the TAMRA quencher. Peptide cleavage efficiency was determined by calculating area under the curve (AUC) with fold increase in FAM florescence over time (time t = 10 to 40 min).

SARS-CoV-2 RNA isolation and sequencing. For the inactivation of SARS-CoV-2, the virus suspension was mixed 1:1 with peqGOLD TriFast (VWR). RNA was then extracted using the innuPREP Virus RNA Kit (Analytik Jena) according to the manufacturer's instructions. The RNA quality was confirmed with Bioanalyzer (Agilent Technologies) and Qubit (Thermo Fisher Scientific). The sequencing libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs) including ERCC RNA Spike-in control. The sequencing was performed with NovaSeq 6000 SP reagent kit (100 cycles) on NovaSeq 6000 System (Illumina).

Raw fastq files were processed to trim adaptor and low-quality reads (fastp 0.20.0). The resulting sequence reads were aligned with reference (BWA 0.7.8). The alignment file was sorted (SAMtools version 1.10) and realigned (GATK IndelRealigner version 3.7). Finally, a pileup file (SAMtools) was created and mutations were detected with VarScan (version 2.3.9). Mutation frequency for whole genome was calculated by extracting total read depth and reference base reads for each nucleotide from the pileup file using SAMtools, where mutation frequency in the population is given as ratio of reference reads to total reads at the interrogated position. Mutation frequencies were converted to percent values and plotted with ggplot2 (R/Bioconductor 4.0). Alignment files were inspected with IGV (version 2.8.3) and Tablet (version 1.19.09.03) to define if nearby mutations in the furin cleavage site were present on the same read.

Mathematical modeling. The mathematical model to fit the curves for the mutation assumes that the proportion of viruses with a mutation on the considered genetic sequence (SARS-CoV-2 genome positions 23,606, 23,607, and 23,616) at time t depends on the initial proportion of mutated viruses $y(0)$, the mutation rate a and the reverse mutation rate b. This leads to the ordinary differential equation $y' = a(1-y)-by$ with the general solution $y(t) = ce^{-(a+b)t} + \frac{a}{a+b}$. Assuming that there are no mutations at time zero, i.e., $y(0) = 0$, then $c = -\frac{a}{a+b}$ follows. If we start with a mutated strain, i.e., $y(0) = 1$, this implies $c = \frac{b}{a+b}$. The mutation rates a and b of the corresponding non-linear curve are estimated based on the data by the optimization procedure optim provided by the R statistics software, minimizing the sum of absolute errors. R script used is provided in supplemental data.

Quantification and statistical analysis. Kruskal-Wallis with Dunn's comparison and two-way analysis of variance (ANOVA) with Bonferroni post-test were used to test for statistical significance. Where appropriate, two-way ANOVA with Sidak's multiple comparison was used. *P values* < 0.05 were considered significant (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$), ****$P < 0.0001$, $P > 0.05$ not significant (n.s.). For all statistical analyses, the GraphPad Prism 7 software package was used (GraphPad Software).

Data, code, and materials availability. The SARS-CoV-2 deep sequencing data generated during this study are available at NCBI sequence read archive under BioProject number PRJNA650134. The R code used in this study are provided in supplemental data (Data set S1). Further information and requests for resources and reagents can be directed to the corresponding authors. The materials will be made available upon receipt of a material transfer agreement (MTA).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.6 MB.

ACKNOWLEDGMENTS

We thank HZI Genome Analytics team for support, especially Michael Jarek. We kindly acknowledge Susanne Talay for helpful discussions and expert technical support. We thank Ayse Barut and Inge Hollatz-Rangosch for technical assistance. This work was supported by the Helmholtz Association through the Helmholtz EU partnering grant PIE-0008 and by the grant 14-76103-84 from the Ministry of Science and Culture of Lower Saxony to LCS.

Conceptualization, M.Z.C. and K.E.; Methodology, M.Z.C., K.E., M.H., M.G., and F.K.; Software and Formal Analysis, M.Z.C., K.E., M.H., and F.K.; Investigation, M.Z.C., K.E., M.H., M.G., L.A., and Y.K.; Data Curation, M.Z.C.; Writing – Original Draft, M.Z.C., K.E., and L.C.-S.; Writing – Review & Editing, M.Z.C., K.E., L.B., S.L., A.K., and L.C.-S.; Supervision, A.K., S.P., and L.C.-S.; Resources, L.B. and S.L.; Funding Acquisition, S.P. and L.C.-S.

REFERENCES

1. Wang C, Horby PW, Hayden FG, Gao GF. 2020. A novel coronavirus outbreak of global health concern. Lancet 395:470–473. https://doi.org/10.1016/S0140-6736(20)30185-9.

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, China Novel Coronavirus I, Research T. 2020. A novel coronavirus from patients with

pneumonia in China, 2019. N Engl J Med 382:727–733. https://doi.org/10.1056/NEJMoa2001017.
3. Neuman BW, Adair BD, Yoshioka C, Quispe JD, Orca G, Kuhn P, Milligan RA, Yeager M, Buchmeier MJ. 2006. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. J Virol 80:7918–7928. https://doi.org/10.1128/JVI.00645-06.
4. Tortorici MA, Veesler D. 2019. Structural insights into coronavirus entry. Adv Virus Res 105:93–116. https://doi.org/10.1016/bs.aivir.2019.08.002.
5. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Muller MA, Drosten C, Pohlmann S. 2020. SARS-CoV-2 Cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181:271–280. https://doi.org/10.1016/j.cell.2020.02.052.
6. Li F, Li W, Farzan M, Harrison SC. 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309:1864–1868. https://doi.org/10.1126/science.1116480.
7. Millet JK, Whittaker GR. 2015. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Res 202:120–134. https://doi.org/10.1016/j.virusres.2014.11.021.
8. Klimstra WB, Heidner HW, Johnston RE. 1999. The furin protease cleavage recognition sequence of Sindbis virus PE2 can mediate virion attachment to cell surface heparan sulfate. J Virol 73:6299–6306. https://doi.org/10.1128/JVI.73.8.6299-6306.1999.
9. Nakayama K. 1997. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. Biochem J 327:625–635. https://doi.org/10.1042/bj3270625.
10. de Haan CA, Haijema BJ, Schellen P, Wichgers Schreur P, Te Lintelo E, Vennema H, Rottier PJ. 2008. Cleavage of group 1 coronavirus spike proteins: how furin cleavage is traded off against heparan sulfate binding upon cell culture adaptation. J Virol 82:6078–6083. https://doi.org/10.1128/JVI.00074-08.
11. Cardin AD, Weintraub HJ. 1989. Molecular modeling of protein-glycosaminoglycan interactions. Arteriosclerosis 9:21–32. https://doi.org/10.1161/01.atv.9.1.21.
12. de Haan CA, Li Z, Te Lintelo E, Bosch BJ, Haijema BJ, Rottier PJ. 2005. Murine coronavirus with an extended host range uses heparan sulfate as an entry receptor. J Virol 79:14451–14456. https://doi.org/10.1128/JVI.79.22.14451-14456.2005.
13. Lau SY, Wang P, Mok BW, Zhang AJ, Chu H, Lee AC, Deng S, Chen P, Chan KH, Song W, Chen Z, To KK, Chan JF, Yuen KY, Chen H. 2020. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. Emerg Microbes Infect 9:837–842. https://doi.org/10.1080/22221751.2020.1756700.
14. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CCS, Boshier FAT, Ortiz AT, Balloux F. 2020. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol 83:104351. https://doi.org/10.1016/j.meegid.2020.104351.
15. Leung K, Shum MH, Leung GM, Lam TT, Wu JT. 2021. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro Surveill 26.
16. Cherian S, Potdar V, Jadhav S, Yadav P, Gupta N, Das M, Rakshit P, Singh S, Abraham P, Panda S. 2021. Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. bioRxiv. https://doi.org/10.1101/2021.04.22.440932:2021.04.22.440932.
17. Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, Guo D, Hu W, Yang J, Tang Z, Wu H, Lin Y, Zhang M, Zhang Q, Shi M, Liu Y, Zhou Y, Lan K, Chen Y. 2020. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerg Microbes Infect 9:761–770. https://doi.org/10.1080/22221751.2020.1747363.
18. Chiem K, Morales Vasquez D, Park JG, Platt RN, Anderson T, Walter MR, Kobie JJ, Ye C, Martinez-Sobrido L. 2021. Generation and characterization of recombinant SARS-CoV-2 expressing reporter genes. J Virol 95. https://doi.org/10.1128/JVI.02209-20.
19. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom K, Zambon M, Ellis J, Lewis PA, Hiscox JA, Matthews DA. 2020. Characterisation of the transcriptome and proteome of SARS-CoV-2 using direct RNA sequencing and tandem mass spectrometry reveals evidence for a cell passage induced in-frame deletion in the spike glycoprotein that removes the furin-like cleavage site. Genome Med 12:68. https://doi.org/10.1101/2020.03.22.002204%.
20. Hoffmann M, Kleine-Weber H, Pohlmann S. 2020. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Mol Cell 78:779–784. https://doi.org/10.1016/j.molcel.2020.04.022.
21. Tay FP, Huang M, Wang L, Yamada Y, Liu DX. 2012. Characterization of cellular furin content as a potential factor determining the susceptibility of cultured human and animal cells to coronavirus infectious bronchitis virus infection. Virology 433:421–430. https://doi.org/10.1016/j.virol.2012.08.037.
22. Andino R, Domingo E. 2015. Viral quasispecies. Virology 479–480:46–51. https://doi.org/10.1016/j.virol.2015.03.022.
23. Lauring AS, Andino R. 2010. Quasispecies theory and the behavior of RNA viruses. PLoS Pathog 6:e1001005. https://doi.org/10.1371/journal.ppat.1001005.
24. Borderia AV, Rozen-Gagnon K, Vignuzzi M. 2016. Fidelity variants and RNA quasispecies. Curr Top Microbiol Immunol 392:303–322. https://doi.org/10.1007/82_2015_483.
25. Cagno V, Tseligka ED, Jones ST, Tapparel C. 2019. Heparan sulfate proteoglycans and viral attachment: true receptors or adaptation bias? Viruses 11. https://doi.org/10.3390/v11070596.
26. Kim SY, Jin W, Sood A, Montgomery DW, Grant OC, Fuster MM, Fu L, Dordick JS, Woods RJ, Zhang F, Linhardt RJ. 2020. Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. Antiviral Res 181:104873. https://doi.org/10.1016/j.antiviral.2020.104873.
27. Kwon PS, Oh H, Kwon SJ, Jin W, Zhang F, Fraser K, Hong JJ, Linhardt RJ, Dordick JS. 2020. Sulfated polysaccharides effectively inhibit SARS-CoV-2 in vitro. Cell Discov 6:50. https://doi.org/10.1038/s41421-020-00192-8.
28. Park JE, Li K, Barlan A, Fehr AR, Perlman S, McCray PB, Jr, Gallagher T. 2016. Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. Proc Natl Acad Sci U S A 113:12262–12267. https://doi.org/10.1073/pnas.1608147113.
29. Hoffmann M, Arora P, Groß R, Seidel A, Hörnich B, Hahn A, Krüger N, Graichen L, Hofmann-Winkler H, Kempf A, Winkler MS, Schulz S, Jäck H-M, Jahrsdörfer B, Schrezenmeier H, Müller M, Kleger A, Münch J, Pöhlmann S. 2021. SARS-CoV-2 variants B.1.351 and B.1.1.248: Escape from therapeutic antibodies and antibodies induced by infection and vaccination. bioRxiv. https://doi.org/10.1101/2021.02.11.430787:2021.02.11.430787.
30. Garcia-Beltran WF, Lam EC, St Denis K, Nitido AD, Garcia ZH, Hauser BM, Feldman J, Pavlovic MN, Gregory DJ, Poznansky MC, Sigal A, Schmidt AG, Iafrate AJ, Naranbhai V, Balazs AB. 2021. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell 184:2372–2383. https://doi.org/10.1016/j.cell.2021.03.013.
31. Kleine-Weber H, Elzayat MT, Wang L, Graham BS, Muller MA, Drosten C, Pohlmann S, Hoffmann M. 2019. Mutations in the spike protein of middle east respiratory syndrome coronavirus transmitted in Korea increase resistance to antibody-mediated neutralization. J Virol 93. https://doi.org/10.1128/JVI.01381-18.
32. Berger Rentsch M, Zimmer G. 2011. A vesicular stomatitis virus replicon-based bioassay for the rapid and sensitive determination of multi-species type I interferon. PLoS One 6:e25858. https://doi.org/10.1371/journal.pone.0025858.