AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

# The SARS-CoV-2 Transcriptome and the Dynamics of the S Gene Furin Cleavage Site in Primary Human Airway Epithelia

**Wei Zou,[a] Min Xiong,[b] Siyuan Hao,[b] Elizabeth Yan Zhang,[c] Nathalie Baumlin,[d] Michael D. Kim,[d] Matthias Salathe,[d] Ziying Yan,[e] Jianming Qiu[b]**

[a]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA
[b]Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, Kansas, USA
[c]GeneGoCell Inc., San Diego, California, USA
[d]Department of Internal Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA
[e]Department of Anatomy and Cell Biology, University of Iowa, Iowa City, Iowa, USA

Wei Zou and Min Xiong contributed equally. Author order was determined by reverse alphabetical order.

**ABSTRACT** The spike (S) polypeptide of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) consists of the S1 and S2 subunits and is processed by cellular proteases at the S1/S2 boundary that contains a furin cleavage site (FCS), $^{682}$RRAR↓S$^{686}$. Various deletions surrounding the FCS have been identified in patients. When SARS-CoV-2 propagated in Vero cells, it acquired deletions surrounding the FCS. We studied the viral transcriptome in Vero cell-derived SARS-CoV-2-infected primary human airway epithelia (HAE) cultured at an air-liquid interface (ALI) with an emphasis on the viral genome stability of the FCS. While we found overall the viral transcriptome is similar to that generated from infected Vero cells, we identified a high percentage of mutated viral genome and transcripts in HAE-ALI. Two highly frequent deletions were found at the FCS region: a 12 amino acid deletion ($^{678}$TNSPRRAR↓SVAS$^{689}$) that contains the underlined FCS and a 5 amino acid deletion ($^{675}$QTQTN$^{679}$) that is two amino acids upstream of the FCS. Further studies on the dynamics of the FCS deletions in apically released virions from 11 infected HAE-ALI cultures of both healthy and lung disease donors revealed that the selective pressure for the FCS maintains the FCS stably in 9 HAE-ALI cultures but with 2 exceptions, in which the FCS deletions are retained at a high rate of >40% after infection of ≥13 days. Our study presents evidence for the role of unique properties of human airway epithelia in the dynamics of the FCS region during infection of human airways, which is likely donor dependent.

**IMPORTANCE** Polarized human airway epithelia at an air-liquid interface (HAE-ALI) are an *in vitro* model that supports efficient infection of SARS-CoV-2. The spike (S) protein of SARS-CoV-2 contains a furin cleavage site (FCS) at the boundary of the S1 and S2 domains which distinguishes it from SARS-CoV. However, FCS deletion mutants have been identified in patients and *in vitro* cell cultures, and how the airway epithelial cells maintain the unique FCS remains unknown. We found that HAE-ALI cultures were capable of suppressing two prevalent FCS deletion mutants (Δ$^{678}$TNSPRRAR↓SVAS$^{689}$ and Δ$^{675}$QTQTN$^{679}$) that were selected during propagation in Vero cells. While such suppression was observed in 9 out of 11 of the tested HAE-ALI cultures derived from independent donors, 2 exceptions that retained a high rate of FCS deletions were also found. Our results present evidence of the donor-dependent properties of human airway epithelia in the evolution of the FCS during infection.

**KEYWORDS** SARS-CoV-2, transcriptome, furin cleavage site, human airway epithelia

**Citation** Zou W, Xiong M, Hao S, Zhang EY, Baumlin N, Kim MD, Salathe M, Yan Z, Qiu J. 2021. The SARS-CoV-2 transcriptome and the dynamics of the S gene furin cleavage site in primary human airway epithelia. mBio 12: e01006-21. https://doi.org/10.1128/mBio.01006-21.

**Editor** Xiang-Jin Meng, Virginia Polytechnic Institute and State University



Address correspondence to Ziying Yan, ziying-yan@uiowa.edu, or Jianming Qiu, jqiu@kumc.edu.

This article is a direct contribution from Jianming Qiu, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Yao-Wei Huang, Zhejiang University, and Zheng Xing, University of Minnesota.

**Received** 6 April 2021
**Accepted** 8 April 2021
**Published** 11 May 2021

The ongoing coronavirus disease 2019 (COVID-19) outbreak, caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), poses a great threat to global public health with a devastating mortality (1–4). The virus has spread with unprecedented speed and has infected >100 million people worldwide, causing >2 million deaths so far. The efficacy of the only U.S. Food and Drug Administration (FDA)-approved drug, Veklury (remdesivir), to treat COVID-19 patients is limited to early phases of the disease and is supportive (5). Ongoing rollout of the adenovirus-based and two mRNA-based COVID-19 vaccines approved by the FDA under an emergency use authorization, with more vaccines to follow soon, is the hope to prevent COVID-19 and contain the virus (6).

SARS-CoV-2 phylogenetically belongs to the genus *Betacoronavirus* of the family *Coronaviridae* (7, 8) and is closely related to the previously identified severe acute respiratory syndrome coronavirus (SARS-CoV) with an identity of 79% in genome sequence (9, 10). SARS had an outbreak in 2002 to 2003 (11, 12). The genome organization of SARS-CoV-2 is the same as other betacoronaviruses. It has six major open reading frames (ORFs) arranged in order from the structured 5′ untranslated region (UTR) to 3′ UTR (13, 14): replicases (ORF1a and ORF1b), spike (S), envelope (E), membrane (M), and nucleocapsid (N). In addition, at least seven ORFs encoding accessory proteins (3a, 6, 7a, 7b, 8, 9a, and 9b) are interspersed between the structural protein genes (10, 15).

The replication and transcription of SARS-CoV-2 largely resemble that of the SARS-CoV (15, 16). The accepted model for coronavirus transcription indicates that all viral mRNAs have a common 5′-leader (L) sequence or the 5′-cap structure at the 5′ UTR, and a common poly(A) tail at the 3′ UTR (15, 17–19). The highly conserved leader sequences contain the transcription regulatory sequences (TRS) which play an important role in viral RNA transcription (18–20). Upon cell entry of the virus, the incoming positive-sense genomic RNA (+gRNA) subjects it to immediate translation of two large ORFs, ORF1a and ORF1b, for viral nonstructural proteins, which form the viral replication and transcription complex (RTC) (21). In the complex, the viral +gRNA serves as the template for the production of negative-sense gRNA (−gRNA) and subgenome RNA (−sgRNA) intermediates, which in turn serve as the templates for the synthesis of +gRNA and +sgRNAs (22). The positive-sense viral RNAs are the mRNAs used for the translations of 16 viral nonstructural proteins, 9 accessory proteins, and 4 structural proteins, which include spike protein S, envelope protein E, membrane protein M, and nucleocapsid protein N (21, 23). Next, the newly synthesized +gRNA is encapsidated by the N protein to assemble progeny virions with other viral structural proteins, M, E, and S (21).

Most SARS-CoV-2 structural and nonstructural proteins share greater than 85% identity in protein sequence with SARS-CoV, whereas their S proteins share an identity of only approximately 77% (2). The S protein consists of two subunits, S1 and S2, and is a key glycoprotein responsible for receptor binding and determining the host tropism, pathogenicity, and transmissibility (24, 25). It forms a homotrimer on the virion surface and triggers viral entry into target cells via binding of the S1 subunit to its cognate receptor, angiotensin-converting enzyme 2 (ACE2) (2, 26, 27). One significant difference among S proteins of SARS-CoV-2, SARS-CoV, and other bat SARS-like coronaviruses, such as bat coronavirus BtCoV-RaTG13, is the addition of 4 amino acids, PRRA, at the S1/S2 boundary (25). This insertion forms a polybasic residue motif, assembling a furin cleavage site (FCS), RRAR↓S, which is highly related to the furin cleavage consensus sequence RX[K/R]R (X, any amino acid) (28). The absence of the FCS in the other betacoronaviruses suggests the insertion of PRRA is a key factor in the virulence of SARS-CoV-2, which has been shown to broaden cell tropism, transmissibility, and pathogenicity of the virus (29–31).

Viral transcriptomes of SARS-CoV-2 have been studied by several groups but only in infected Vero cells (15, 17), which revealed quick mutations in the S1/S2 boundary of the S gene after a few passages, including the loss of the FCS and the immediately

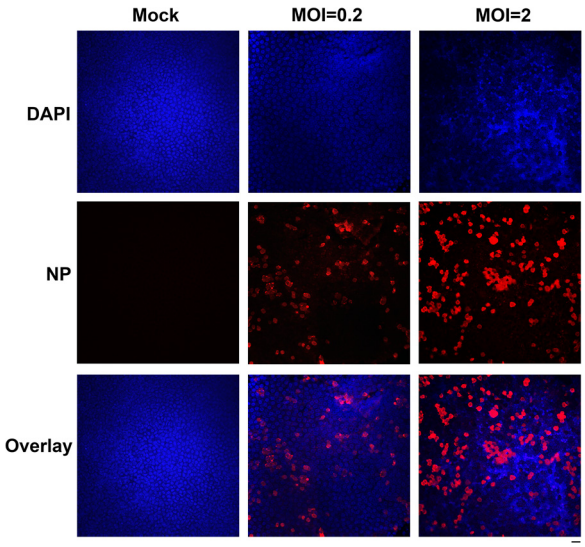


**FIG 1** Immunofluorescence analysis of SARS-CoV-2 infection of HAE cells. HAE-ALI$^{B2\text{-}20}$ cultures were infected with SARS-CoV-2 at an MOI of 0.2 or 2 PFU/cell or, as indicated, mock -infected (Mock). At 4 days postinfection, a piece of the insert membrane was fixed in 4% paraformaldehyde in PBS at 4°C overnight and subjected to direct immunofluorescence analysis. The membranes were stained with anti-SARS-CoV-2 N protein (NP). Images were taken on a Leica TCS SPE confocal microscope under 40×, which was controlled by Leica Application Suite X software. The nuclei were stained with DAPI (4′,6-diamidino-2-phenylindole). Bar, 20 $\mu$m.

adjacent amino acids upstream or downstream of the FCS. The loss of the FCS has been identified in progeny virions replicated in Vero cells (17, 32–35). The mutant viruses were stable, quickly took over the wild-type (WT) virus, and became the dominant population during passaging. Of note, various deletions surrounding the FCS have been identified in patients. This raises the question of how the FCS region deletions are selected in human airways.

In this study, we used transcriptome sequencing (RNA-seq) to analyze the viral transcriptome of SARS-CoV-2 in the infected human airway epithelia (HAE) cultured at an air-liquid interface (HAE-ALI), which mimics natural viral infection of human airways (36, 37). While the viral transcriptome overall recapitulated that in Vero cells, we discovered that there is a selective pressure in HAE-ALI to suppress the deletions at the S1/S2 boundary and that this pressure appears individual donor dependent. We identified two FCS region deletions that are strikingly amplified in two HAE-ALI cultures after 2 to 3 weeks of infection, whereas these deletions were suppressed in nine other HAE-ALI cultures.

## RESULTS

**The SARS-CoV-2 transcriptome in SARS-CoV-2-infected HAE-ALI cultures.** HAE-ALI$^{B2\text{-}20}$ cultures were infected with SARS-CoV-2 at a multiplicity of infection (MOI) of 0.2 or 2 or mock infected. At 4 days postinfection (dpi), immunofluorescence assay for the SARS-CoV-2 N protein expression revealed effective SARS-CoV-2 infection in these cultures, with ~10% and ~30% of cells positive in the infections at MOIs of 0.2 and 2, respectively (Fig. 1). This result was similar to our previous observation (37).

Total RNA samples were extracted from infected HAE-ALI cultures at 4 dpi and subjected to reverse transcription, followed by DNA nanoball sequencing. An average total reads of 18.27% and 26.54% were mapped to the SARS-CoV-2 reference genome (Wuhan-Hu-1 isolate; GenBank accession no. MN908947) in the groups of MOI 0.2 and MOI 2, respectively (Table 1). No significant difference was observed in the total reads in the two groups. Notably, the RNA-seq data obtained from SARS-CoV-2-infected Vero-E6 cells had up to 70% of the reads mapped to the viral genome (15), which was likely due to the high infectivity of Vero-E6 cells and that not all the cell types in HAE-

**TABLE 1** Summary of RNA-seq data of SARS-CoV-2 and mock-infected HAE-ALI cultures

| Sample | No. of total reads | No. of mapped viral reads | Mapped viral reads (%) |
|---|---|---|---|
| Mock-1 | 44,690,289 | 0 | 0 |
| Mock-2 | 44,710,969 | 0 | 0 |
| Mock-3 | 44,661,039 | 0 | 0 |
| MOI0.2-1 | 44,665,846 | 9,109,261 | 20.39 |
| MOI0.2-2 | 44,606,709 | 9,446,219 | 21.18 |
| MOI0.2-3 | 44,625,565 | 5,907,458 | 13.24 |
| MOI2-1 | 44,622,724 | 14,536,455 | 32.58 |
| MOI2-2 | 44,658,564 | 12,466,502 | 27.92 |
| MOI2-3 | 44,616,157 | 8,529,469 | 19.12 |

ALI are permissive to the infection (37). Also, for the whole viral genome coverage, in contrast to the observation in SARS-CoV-2-infected Vero-E6 cells (15), we did not observe an obvious 5′-leader peak in the infected HAE cells (Fig. 2). Instead, we observed ~2-fold higher reads in the 3′ end than that in the 5′-end viral genome.

We further analyzed the viral sgRNA expression in infected HAE cells. Junction-spanning reads covering the 5′ leader and different sgRNAs were counted and analyzed (see Data Set S1 in the supplemental material). Different sgRNAs were abundantly expressed in infected HAE cells. As the negative-strand intermediates are only ~1% as abundant as their positive-sense counterparts (22, 38), this indicates most of the identified sgRNAs were +sgRNAs. N-protein-encoding RNA was the most abundantly expressed viral transcript and accounted for 23.11% and 16.93% of total junction-spanning reads in the groups of MOI 0.2 and MOI 2, respectively, followed by ORF3a, ORF7a, M, ORF8, S, E, ORF6 coding RNAs (Fig. 3). The junction-spanning reads associated with ORF7b and ORF9a/b were identified at a level of 0.01% or less of the total junction-spanning reads and were identified in only part of all the six samples in two MOI groups (Data Set S1). In SARS-CoV-2-infected HAE cells, S RNA transcript was expressed at a ratio of ~2% of total junction-spanning reads in both groups (Fig. 3), compared to that of ~8% in Vero cells. We detected relatively higher level of ORF3a (~8%) transcript in SARS-CoV-2-infected HAE cells (Fig. 3), in contrast to 5.22% in infected Vero cells (15).

Interestingly, in all identified spanning-junction reads, only ~50% correlated with the canonical sgRNA transcripts in both MOI infection groups. The other half junction-spanning reads represent either reads covering 5′-leader sequence but with unex-

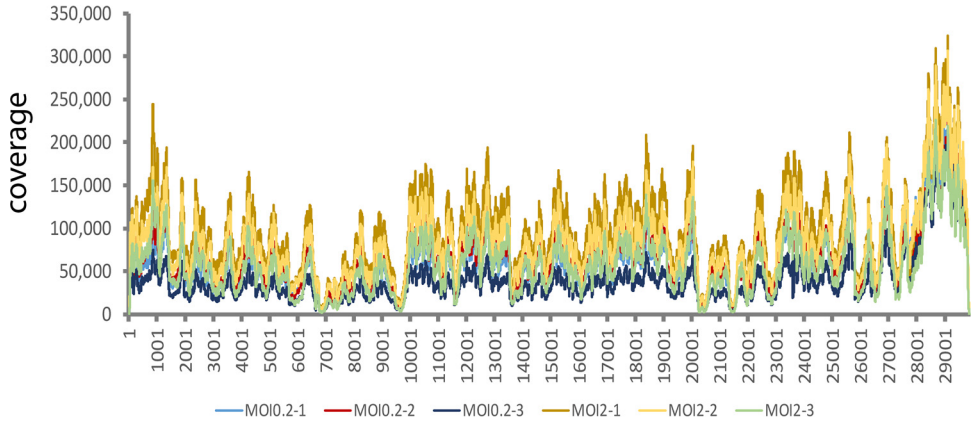


**FIG 2** Genome coverage of SARS-CoV-2-infected HAE cells with MOIs of 0.2 and 2, respectively. Six total RNA samples, as indicated by six colors, extracted from HAE-ALI$^{B2\text{-}20}$ cultures infected with SARS-CoV-2 at MOIs of 0.2 and 2, respectively, were subjected to whole RNA-seq. The reads were mapped to the reference SARS-CoV-2 Wuhan-Hu-1 strain genome (GenBank accession no. MN908947, NCBI), as shown with nucleotide numbers (*x* axis), using BWA, and the sequencing read coverage (*y* axis) was calculated.

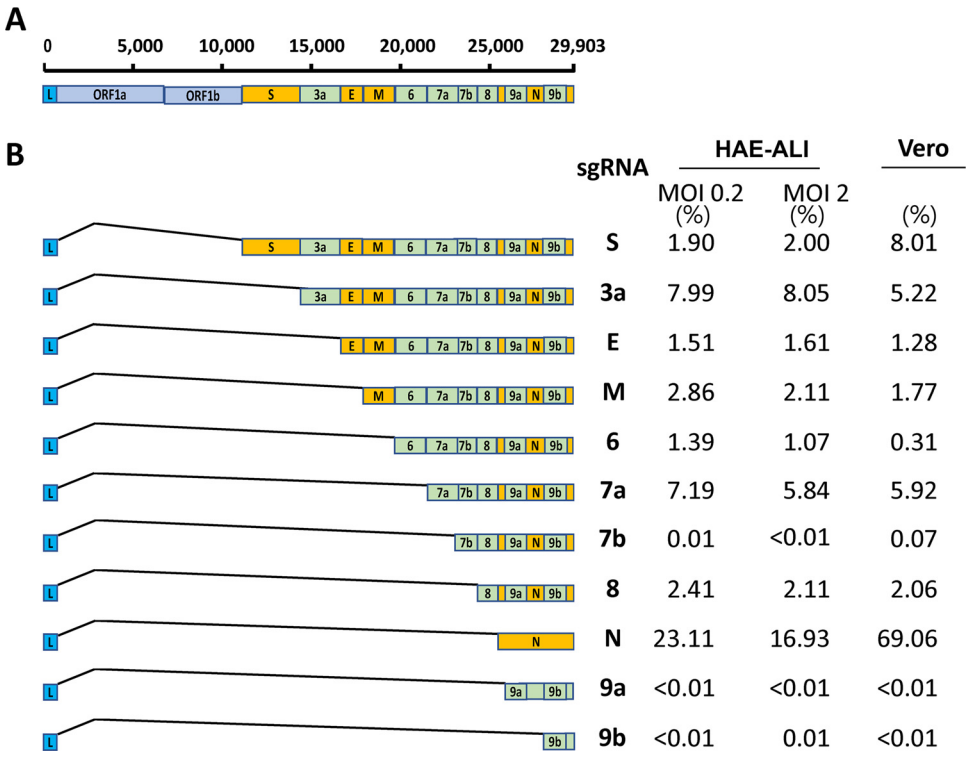


**FIG 3** Identification and quantification of SARS-CoV-2 subgenomic RNAs. (A) Genome organization. The SARS-CoV-2 genome is schematically diagrammed (not to scale) with regions in order coding for open reading frame 1a (ORF1a)/ORF1b, S protein, ORF3a, E and M proteins, ORF7a/b and ORF8, N protein, and ORF9a/b. The leader sequence was labeled L in a blue box. The structural genes are labeled within boxes in orange, and the accessory genes are labeled within boxes in light green. (B) Subgenomic RNAs. Six total RNA samples were extracted from SARS-CoV-2-infected HAE-ALI cultures (at MOIs of 0.2 and 2, respectively) and subjected to whole RNA-seq. Three repeats in each MOI group were merged. Junction-spanning reads were identified using STAR (2.7.3a), and the transcript abundance, as shown as a percentage under HAE-ALI/MOI of 0.2 or 2, was estimated by counting the reads that span the junction of the corresponding RNA transcript. The left is the diagrammed subgenomic RNAs. The canonical junction-spanning reads related to each sgRNA were calculated, and the ratios are shown on right. The abundances of the subgenomic transcripts identified in Vero cells in a previous study (15) are listed for comparison.

pected 3′ sites located in the middle of annotated ORFs or reads covering between different ORFs or inside an ORF without 5′-leader sequence (Data Set S1). It is important to note that a lot of these noncanonical junction-spanning patterns were supported by only one read from the RNA-seq data, indicating that these noncanonical transcripts may arise from erroneous replicase activity.

**Identification of deletions surrounding the furin cleavage site at S1/S2.** Among the ~50% noncanonical junction-spanning reads, we identified a high abundant 36-bp deletion, mut-del1, located at nucleotide (nt) 23,594 to 23,629 spanning the FCS (Fig. 4A) that encodes amino acids (aa) $^{678}$TNSPRRAR↓SVAS$^{689}$ (Fig. 4B, "↓" indicates cleavage). It displayed at frequencies of 21.04% and 14.79% of total junction-spanning reads in MOI 0.2 and MOI 2 groups, respectively (Table 2). Another 15-bp deletion, mut-del2, located at nt 23,583 to 23,597 and encoding aa $^{675}$QTQTN$^{679}$ (Fig. 4A), just two amino acids ahead of the FCS, was also identified. It accounted for 0.42% and 15.11% of the total junction-spanning reads in MOI 0.2 and MOI 2 groups, respectively (Table 2). The ratio of mut-del1 is only slightly lower than the N sgRNA and nearly 10 times higher than the S sgRNA transcript, indicating a high fraction of this mutation comes from the viral genome (+gRNA).

To further reveal the ratio of the two deletions in total viral genome, the junction-spanning reads associated with the two deletions were normalized with the average reads covering the same deletions. The results showed that 26.31% and 6.67% of the

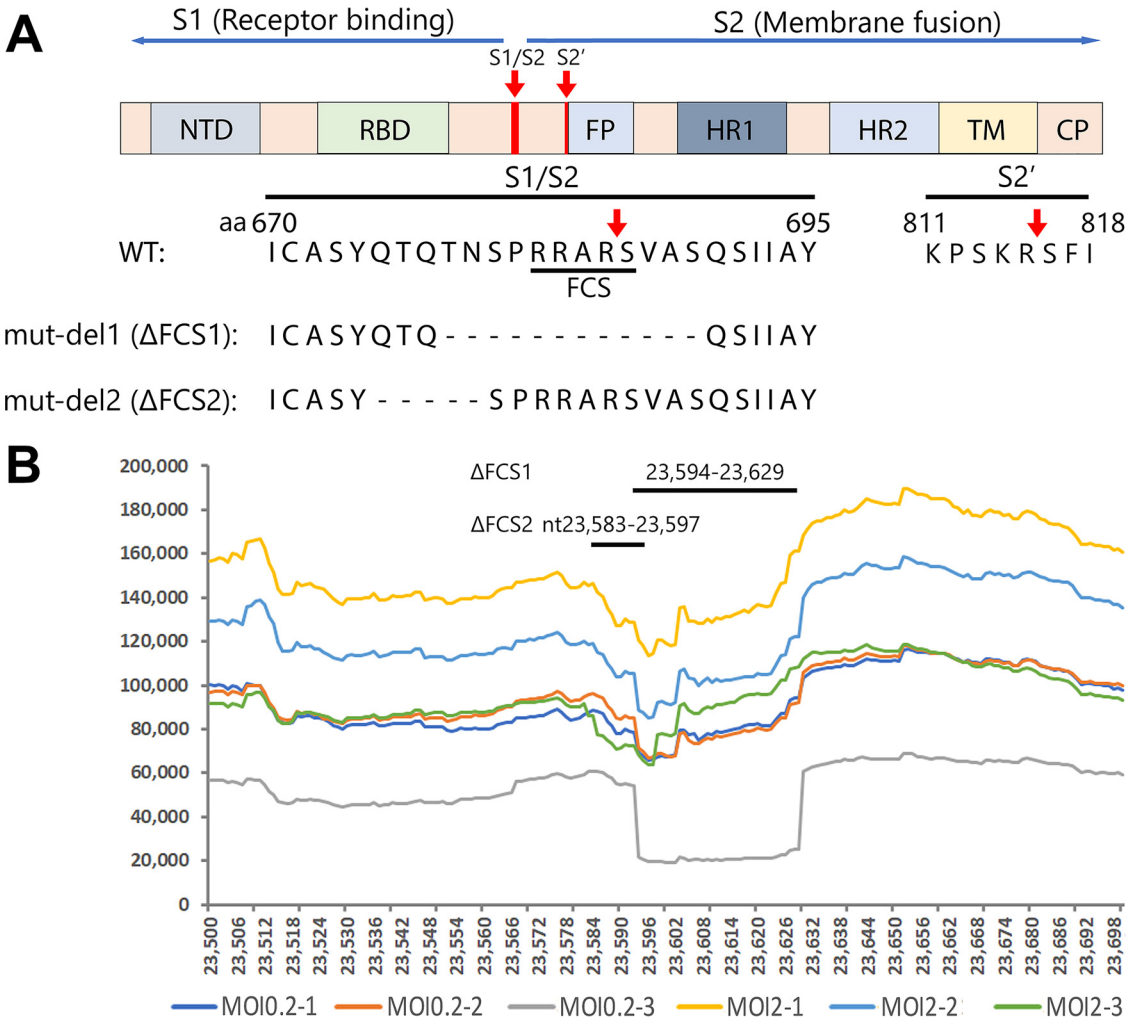


**FIG 4** Features of the S gene of SARS-CoV-2 and the deletions detected in the FCS region. (A) S gene and FCS. Key domains of the S polypeptide are diagrammed in the context of the SARS-CoV-2 genome. The S1 protein, receptor binding unit, harbors N-terminal domain (NTD) and receptor binding domain (RBD) subunit, which is conserved and recognizes ACE2. The S2, membrane fusion subunit, has fusion peptide (FP), S2' proteolytic site, two heptad repeats, HR1 and HR2, and a transmembrane domain (TM) followed by cytoplasmic peptide (CP) (31). The S protein has acquired a polybasic site (RRAR↓S, a furin cleavage site [FCS]) for cleavage at the S1/S2 boundary. An FCS region of aa 670 to 695, together with the two key deletions mut-del1 (ΔFCS1) and mut-del2 (ΔFCS2), are shown with S amino acid sequences of the SARS-CoV-2 genome (GenBank accession no. MN908947). (B) Coverage plots of S gene at nt 23,500 to 23,698 in SARS-CoV-2-infected HAE-ALI$^{B2\text{-}20}$. The coverage plots show that the most abundant junction-spanning reads in SARS-CoV-2-infected HAE-ALI$^{B2\text{-}20}$ cultures are the 36-bp and 15-bp deletions in the S gene of nt 23,594 to 23,629 and nt 23,583 to 23,597, respectively, which deleted 12 aa and 5 aa shown in mut-del1 and mut-del2 in panel A.

viral reads related to this region contain the mut-del1 deletion, while 0.37% and 8.17% of this region contain mut-del2 deletion in MOI 0.2 and MOI 2 groups, respectively (Table 2). It should be noted that the total reads used for normalization include reads of both viral gRNA and sgRNAs. Thus, here we were unable to distinguish the origin of these two deletions from the viral genome and viral RNA transcripts in these total cellular transcriptome data.

Except for these two highly abundant mut-del1 and mut-del2 deletions, we also observed a 21-bp FCS deletion at nt 23,595 to 23,615, encoding aa $^{678}$TNSPRRA$^{684}$, but only in the MOI 2 group with 1.13% of the total junction-spanning reads, and a 39-bp deletion at the N terminus of the S protein (nt 21,743 to 21,781 encoding aa $^{61}$NVTWFHAIHVSGT$^{73}$) with 0.27% and 0.60% of the total junction-spanning reads in MOI 0.2 and MOI 2 groups, respectively.

In addition to these deletions in the S gene, we identified about 50 different in-

**TABLE 2** Ratio of reads covering mut-del1 and mut-del2 to total junction spanning reads and viral genome

| | Junction-spanning reads (%)[a] | | Viral genome ratio (%)[b] | |
|---|---|---|---|---|
| Mutant | MOI 0.2 | MOI 2 | MOI 0.2 | MOI 2 |
| Mut-del1 | 21.04 | 14.79 | 26.31 | 6.67 |
| Mut-del2 | 0.42 | 15.11 | 0.37 | 8.17 |

[a]The minimal size of the junctions was set at 10 as described in Materials and Methods.
[b]The total reads include both viral genome RNA (gRNA) and subgenomic RNA (sgRNA).

frame or frameshift deletions in the M encoding region that appeared in all six samples of both MOI groups, and there were even more deletions in the M coding region that appeared in only a part of the six RNA samples (Data Set S1). Although the ratio of single deletion was low, the 50 deletion patterns that appeared in all six RNA samples had the ratios of 2.39% and 3.18% in MOI 0.2 and MOI 2 groups, respectively, which is similar or even higher than the identified canonical junction-spanning reads related to M sgRNAs (Fig. 3). Notably, most of these identified deletion patterns of the M gene also appeared in SARS-CoV-2-infected Vero cells (15). Whether these deletions produce functional M protein or affect the function of M protein warrants further studies. In SARS-CoV-2-infected Vero-E6 cells, a high ratio of 27-bp deletion in E gene (nt 26,257 to 26,283) was identified (15), which, however, was not found in infected HAE cells.

**Dynamics of the FCS region deletions in virions apically released from SARS-CoV-2-infected HAE-ALI cultures derived from various donors.** To further investigate the FCS region deletions during SARS-CoV-2 infection of HAE cells, we infected HAE-ALI cultures generated from five different donors, B3-20 (MOI = 0.2), B4-20 (MOI = 0.2 and 2), B9-20 (MOI = 2), L209 (MOI = 0.2), and KC19 (MOI = 0.2), and collected the progeny in the apical washes at different time points. The dynamics of apical virus release of the HAE-ALI cultures of B3-20, B4-20, B9-20, and L209 have been described in our previous study (37). The apical virus release kinetics of the HAE-ALI$^{KC19}$ is shown in Fig. 5A. Viral RNA was prepared either for RNA-seq or for PCR amplicon-seq of a 384-nt sequence covering the FCS. Notably, mut-del1 was not significantly detected ($<$0.1%) in all the apically released viruses collected at $>$13 dpi (Table 3, Bx-20). Nevertheless, for viruses collected from HAE-ALI$^{KC19}$, the mut-del2 was detected at a high level (20.75%$^{RNA\text{-}seq}$ and 20.98%$^{RNA\text{-}seq}$) at 4 dpi and 13 dpi, respectively, which reached a close level of 41.79%$^{PCR\text{-}seq}$ at 21 dpi. Although the viruses derived from HAE-ALI$^{B3\text{-}20}$, HAE-ALI$^{L209}$, and HAE-ALI$^{B4\text{-}20}$ contain a high level (23.17%, 30.33%, and 8.3%, respectively) of mut-del2 at 3 dpi, it decreased to a level of $<$2% at $\geq$17 dpi (Table 3, Bx-20). HAE-ALI$^{B9\text{-}20}$ never produced significant mut-del2 ($<$0.1%).

Notably, SARS-CoV-2 isolate USA-WA1/2020 P0 stock provided by BEI was already passaged four times in Vero cells, and it was reported that there appeared significant heterogeneity at the S1/S2 boundary in Vero cell-propagated virus (32). To verify this, we sequenced the viral RNA of passage 0 (P0) (the originally received vial) and P1 (passaged once in Vero-E6 cells) virus stocks. The results showed that there was no detectable mut-del1 in the P0 stock but a high rate of 21.69%$^{RNA\text{-}seq}$ and 40.47$^{PCR\text{-}seq}$ in the P1 stock. However, while there was ~2% of mut-del2 in the P0 stock, it slightly increased to only ~5% in the P1 stock (Table 3, P0 and P1). These results confirmed that even though there was no or a low level of mut-del1 and mut-del2 in the P0 stock, there was a high level of mut-del1 and a low level of mut-del2 in the P1 stock, which we used for infection of all HAE-ALI cultures.

Taken together, the results demonstrated that the mut-del1 appears at a very low rate in all the HAE-ALI produced viruses at late time points of infection ($\geq$17 dpi), which was confirmed by both RNA-seq and PCR amplicon-seq. Although the inoculum had the mut-del1 detected at a high frequency rate of 21.69%$^{RNA\text{-}seq}$ (40.47%$^{PCR\text{-}seq}$), this deletion was obviously suppressed when the virus propagated in the HAE-ALI cultures. In addition, except for the viruses produced from HAE-ALI$^{KC19}$, mut-del2 also

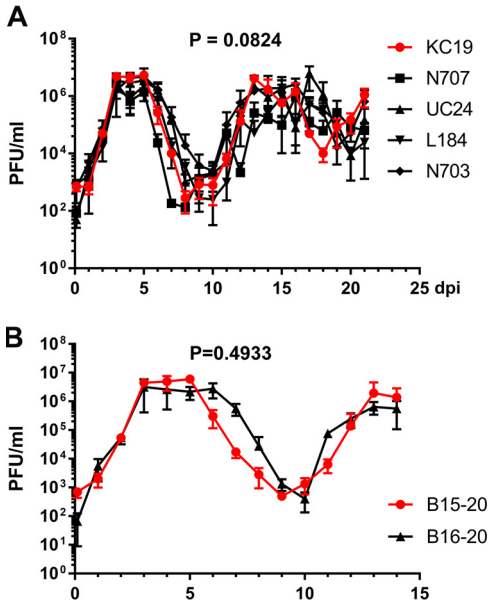


**FIG 5** Apical virus release kinetics of SARS-CoV-2-infected HAE-ALI cultures. (A and B) HAE-ALI$^{KC19}$ and four COPD HAE-ALI cultures (HAE-ALI$^{N707}$, HAE-ALI$^{UC24}$, HAE-ALI$^{L184}$, and HAE-ALI$^{N703}$) (A) and HAE-ALI$^{B15\text{-}20}$ and HAE-ALI$^{B16\text{-}20}$ (B) cultures were infected with SARS-CoV-2 at an MOI of 0.2 from the apical side. At the indicated days postinfection (dpi), the apical surface was washed with 300 $\mu$l of D-PBS to collect the released viruses. PFU were determined (*y* axis) and plotted to the dpi. Values represent means ± standard deviations (SD) (error bars). The graphs in panels A and B were prepared with GraphPad Prism 9.0.2.

appeared at a low detection rate during the course of infection (≥17 dpi) in various HAE-ALI cultures.

**FCS region deletions during SARS-CoV-2 infection of human airway epithelia are donor dependent.** The above RNA-seq and PCR-seq results of apically released vi-

**TABLE 3** Summary of the detections of mut-del1 and mut-del2 in stock viruses and apical washes of SARS-CoV-2-infected HAE-ALI cultures derived from different donors

| | | Mut-del1 | | Mut-del2 | |
|---|---|---|---|---|---|
| Virus or donor | Source or dpi | RNA-seq (%) | PCR-seq (%) | RNA-seq (%) | PCR-seq (%) |
| P0 stock | BEI | 0.87 | 2.03 | 2.09 | 2.45 |
| P1 stock | Vero-E6 | 21.69 | 40.47 | 5.18 | 5.16 |
| B3-20 (0.2[c]) | 3 | ND[a] | | 23.17 | |
| | 12 | ND | | 4.44 | |
| | 20 | | 0.00 | | 0.57 |
| B4-20 (2) | 3 | 1.08 | | 8.33 | |
| | 12 | | 0.052 | | 4.27 |
| | 13 | | 0.033 | | 3.90 |
| | 17 | | 0.103 | | 1.59 |
| B4-20 (0.2) | 3 | | 1.58 | | 1.48 |
| | 14 | | 0.07 | | 1.41 |
| B9-20 (2) | 5 | | 0.001 | | 0.01 |
| | 11 | | 0.064 | | 0.09 |
| | 17[b] | | 0.056 | | 0.15 |
| | 17[b] | | 0.024 | | 0.07 |
| KC19 (0.2) | 4 | 0.16 | 1.79 | 20.75 | 35.71 |
| | 13 | ND | | 20.98 | |
| | 21 | | 0.017 | | 41.79 |
| L209 (0.2) | 3 | | 0.037 | | 30.33 |
| | 14 | | 0.001 | | 0.11 |
| | 41 | | 0.001 | | 0.04 |

[a]ND, not detected or <0.1%.
[b]Independent samples.
[c]Numbers in parentheses are MOIs.

**TABLE 4** Detections of mut-del1 and mut-del2 in SARS-CoV-2 virions apically released from infected HAE-ALI cultures derived from B15-20 and B16-20 donors (MOI = 0.2)

| Donor | dpi | PCR-seq (%) of mut-del1 | PCR-seq (%) of mut-del2 |
|---|---|---|---|
| B15-20 | 3 | 31.87 | 0.18 |
| | 13 | 54.22 | 0.69 |
| B16-20 | 3 | 6.78 | 0.71 |
| | 13 | 0.08 | 0.16 |

rions from five individual HAE-ALI cultures suggested a selective pressure in suppressing the deletions of the FCS region during SARS-CoV-2 propagation in human airway epithelia. The exception is the infection in HAE-ALI$^{KC19}$ cultures, which amplified the mut-del2 to a high level. To address the possibility of the donor dependency of the FCS region deletions, we infected HAE-ALI cultures generated from two additional donors, B15-20 and B16-20, and collected both the viral progeny in apical washes at the early and late time points postinfection for PCR-seq. The apical washes from infected HAE-ALI$^{B15-20}$ and HAE-ALI$^{B16-20}$ had similar kinetics of apical virus release, reaching titers of $>10^6$ PFU/ml at both 3 dpi and 13 dpi (Fig. 5B), indicating that the virus replicated in these two HAE-ALI cultures at a similar level. The sequencing results of the apically released viruses from infected HAE-ALI$^{B15-20}$ showed that mut-del1 was detected at a rate of 31.87% at 3 dpi and an increased level of 54.22% at 13 dpi (Table 4, B15-20/mut-del1). However, mut-del2 was barely detectable at a low rate (0.18%) at 3 dpi, and this rate remained very low at 0.69% at 13 dpi (Table 4, B15-20/mut-del2). For the viruses apically released from infected HAE-ALI$^{B16-20}$ cultures, a stark difference was that the mut-del1 was suppressed during the infection period, whereas mut-del2 was also detected at very low levels ($<1\%$) at both 3 and 13 dpi (Table 4, B16-20). The mut-del1 was detected at a rate of 6.78% at 3 dpi and nearly disappeared (at a rate of 0.08%) by the end of infection (13 dpi). The suppression of mut-del1 in HAE-ALI$^{B16-20}$ cultures was similar to what was observed in the previously tested five ALI cultures (Table 3).

Now that we have tested seven HAE-ALI cultures from seven healthy donors, we next looked into the FCS deletions in four infected HAE-ALI cultures derived from bronchiolar epithelial cells isolated from four patients who had chronic obstructive pulmonary disease (COPD). While we did not observe significant differences in the titers of apically released viruses over the course of 21 days (Fig. 5A), we found that the FCS was largely retained at a rate of $>99\%$ in HAE-ALI$^{N707}$, HAE-ALI$^{UC24}$, and HAE-ALI$^{L184}$ cultures at 3 dpi and in HAE-ALI$^{N703}$ culture at 21 dpi (Table 5).

Together with the detection rates of the viruses produced from the infected HAE$^{KC19}$-ALI cultures, the above results suggested that the probability of FCS region deletions is dependent on the HAE-ALI cultures made from airway epithelial cells of different donors. In contrast to the HAE-ALI$^{KC19}$ that produced a high rate of mut-del2, HAE-ALI$^{B15-20}$ tended to generate the viruses that have a high rate of the mut-del1 deletion.

## DISCUSSION

In this study, we analyzed the transcriptome of SARS-CoV-2 in polarized human bronchial airway epithelia, an *in vitro* model mimicking the SARS-CoV-2 infection in human lower airways (36, 37). We found that the transcriptome in HAE-ALI reflects more closely the viral transcriptome in the airways of COVID-19 patients, providing further support for HAE-ALI as a physiologically relevant *in vitro* culture model to study SARS-CoV-2. Neither RNA-seq data of clinical SARS-CoV-2-positive nasopharyngeal specimens nor RNA-seq of SARS-CoV-2-infected HAE-ALI showed the 5′-leader sequence read peak (39, 40). In SARS-CoV-2-infected Vero-E6 cells, N sgRNAs accounted for up to 69% in total viral RNA transcripts, and S sgRNAs accounted for 8% of the total junction-spanning reads (15) (Fig. 3). Nevertheless, in HAE-ALI cultures,

**TABLE 5** Summary of the detections of mut-del1 and mut-del2 in apical washes of SARS-CoV-2 infected HAE-ALI cultures derived from four COPD donors (MOI = 0.2)

| | | PCR-seq (%) | |
|---|---|---|---|
| **COPD donor** | **dpi** | **Mut-del1** | **Mut-del2** |
| N707 | 3 | 0.171 | 0.75 |
| | 13 | 0.001 | 0.05 |
| | 21 | 0.003 | 0.05 |
| UC24 | 3 | 0.345 | 0.47 |
| | 13 | 0.007 | 0.07 |
| L184 | 3 | 0.327 | 1.11 |
| | 14 | 0.002 | 0.40 |
| | 21 | 0.004 | 0.68 |
| N703 | 3 | 0.602 | 7.15 |
| | 13 | 0.195 | 3.33 |
| | 21 | 0.002 | 0.47 |

SARS-CoV-2 still expresses the abundant N protein transcript and a relatively low level of S gene mRNA. Of note, the overall sgRNA transcripts in infected HAE-ALI were mapped to only 50% of all the canonical sgRNAs, much lower than that in Vero cells (15), which is partially due to the high deletion rate of the FCS region derived from the inoculated virus (P1 stock, Table 3). Except for those arising from erroneous replicase events, some of these noncanonical transcripts may play unknown but functional roles in the coronavirus life cycle.

Among all the SARS-CoV-2 viral genes, the S gene is the most variable one, in particular the S1/S2 junctional region which features the FCS. Increasing evidence has shown that the S1/S2 FCS region is highly unstable, and various deletions and mutations have been detected or isolated in SARS-CoV-2-infected Vero cells (17, 32–35). A mutant with a 30-bp deletion, encoding aa $^{679}$NSPRRAR↓SVA$^{688}$, showed enhanced replication ability in Vero cells and had the capability to dominate the genome population during passage in Vero cells (35). A 21-bp deletion encoding aa $^{679}$NSPRRAR$^{686}$ was detected (>10%) in low (<2 to 3) passaged isolates (33). However, detection of the original clinical specimen where the mutant was derived and SARS-CoV-2-positive clinical specimens showed no such deletions (15 to 30 bp) in the FCS region (34), indicating that the mut-del1 or mut-del1-like (containing FCS) deletions are generated during the propagation in Vero cells. Apparently, SARS-CoV-2 is under strong selection pressure in Vero cells to acquire adaptive mutations in the S protein. Nevertheless, mut-del2 ($^{675}$QTQTN$^{679}$) has been identified not only in Vero cell-passaged isolates (41) but also in 3 of 68 clinical specimens (33), indicating that mut-del2 may be clinically more important (relevant) than mut-del1.

The S protein as a part of the viral envelope facilitates viral entry into infected cells. The S1 subunit contains the receptor binding domain, and the S2 domain mediates fusion of the viral envelope with a cellular membrane (24). The infectivity of SARS-CoV-2 necessitates the activation of S protein. There are two proteolytic activation events associated with S-mediated receptor binding and membrane fusion. The first is a priming cleavage that occurs at the S1/S2 boundary, and the second is the obligatory triggering cleavage that occurs within the S2′ site (Fig. 4A). The priming cleavage at the S1/S2 boundary causes the conformation changes of the S1 subunit for receptor binding and of the S2 subunit for conversion of a fusion competent form, by enabling the S protein to better bind receptors or expose the hidden S2′ cleavage site. The cleavage at S2′ triggers the fusion of the viral envelope with the host cell membrane (24). Cleavage by furin at the S1/S2 site is required for subsequent transmembrane serine protease 2 (TMPRSS2)-mediated cleavage at the S2′ site during viral entry into lung cells (27). However, a cathepsin B/L-dependent auxiliary activation pathway is available during infection of SARS-CoV-2 infection in TMPRSS2-negative cells (35, 42), which is

likely not dependent on the cleavage at S1/S2 (43). One important novel finding of our study is that HAE-ALI cultures prepared from human airway cells isolated from different donors selected different FCS deletions. While most (9/11; 81.8%) of the HAE cultures (B3-20, B9-20, L209, B16-20, and 4 COPD) strongly selected the FCS during virus replication after a long-term infection (the FCS deletions accounted for only $<$1% at 13 dpi), HAE-ALI$^{KC19}$ preferred selection of mut-del2 ($^{675}$QTQTN$^{679}$) (41.79%$^{PCR\text{-}seq}$ at 21 dpi), and HAE-ALI$^{B15\text{-}20}$ selected the mut-del1 ($^{678}$TNSPRRAR↓SVAS$^{689}$) at a rate of 54.22% $^{PCR\text{-}seq}$ at 13 dpi. Although mut-del2 retains the FCS, deletion of QTQTN upstream of the FCS also prevented the cleavage (41). These mutants with amino acid deletions immediately upstream of FCS, like mut-del2, or downstream ($^{685}$RSV$^{687}$ or $^{689}$SQS$^{691}$) also showed significant defects in S protein processing (41, 42). Both types of the FCS region deletions were unable to utilize the furin and TMPRSS2-mediated plasma membrane fusion entry pathway and exhibited a more limited range of cell tropism (42, 44), which might be randomly selected. This is substantiated by the fact that there were no FCS region deletions detected in SARS-CoV-2 propagated in TMPRSS2-expressing cells (42).

Overall, we believe that human airway epithelial cells express ACE2 and TMPRSS2 (37, 45–47), which plays an important role during S protein priming and viral entry, and the virus entry is mediated by the membrane fusion pathway. However, from 2 out of the 11 HAE-ALI cultures tested in this study, the lack of suppression of the FCS region deletion was also found in apically released virions of the infected HAE-ALI cultures made from KC19 and B15-20 donors. Previously, we discovered that the SARS-CoV-2 infection in HAE-ALI resulted in periodic recurrent replication peaks of progeny (37). Since the cleavage at the S2′ site by TMPRSS2 necessitates the priming cleavage at S1/S2, the accumulation of FCS mutations in the progeny during the infection in HAE-ALI$^{B15\text{-}20}$ and HAE-ALI$^{KC19}$ may limit infection of the mut-del1 or mut-del2. Since the progeny virus was a pool mixed with WT and mutants, we did not observe a significant difference in apical virus release in HAE-ALI$^{B15\text{-}20}$ and HAE-ALI$^{KC19}$. In a future study, we will plaque purify the two FCS mutants and examine their infectivity in HAE-ALI derived from various donors, as well as their transmissibility and pathogenicity in animals. We speculate that epithelial cells from these two donors may express much less TMPRSS2, and therefore, the virus utilizes the TMPRSS2-independent and cathepsin-dependent endosomal entry pathway (42, 44, 48), which likely does not require the S cleavage at S1/S2 (43) and thus prefers replication of the FCS deletion mutants.

SARS-CoV-2 infection is less studied in primary airway epithelial culture derived from individuals who already have critical pulmonary disease, such as COPD. COPD is a chronic inflammatory lung disease that causes obstructed airflow in the lung (49). The severity of COVID-19 in COPD patients is in general worse (50, 51). Here, we did not observe an increased replication of SARS-CoV-2 in the four COPD HAE-ALI cultures. However, the FCS deletions were strongly suppressed. In the future, it is worthwhile to look into the retention of the FCS site in HAE-ALI cultures derived from donors of different sexes, age groups, and other chronic lung diseases, as well the expression levels of the cellular proteases furin, TMPRSS2, and cathepsin in these HAE-ALI cultures.

Importantly, the deletion of QTQTN (mut-del2) diminished SARS-CoV-2 entry and infection in Vero-E6 cells (41). Furthermore, three FCS-related deletion mutants, ΔPRRA↓, ΔRAR↓SVAS, and ΔNSPRRAR↓SVA, have been shown to have reduced replication *in vitro* and lung disease in animal models (44, 52, 53), strongly supporting that the FCS is a virulence-related motif. Since the ΔQTQTN also abolished furin cleavage (41), we speculate that the mut-del2 mutant should have reduced lung disease in animals as well. Since the FCS is a key motif related to virulence, it is important to investigate the natural occurrence rate of the FCS region deletions, possibility or limitation of their human-to-human transmission, as well as their pathogenicity. Several studies tried to screen the FCS region deletions from patients-derived SARS-CoV-2. As discussed above, screening of 27 SARS-CoV-2-positive clinical specimens, including one specimen that had FCS deletions identified after passaging in Vero-E6 cells, failed to

detect any FCS deletions (34). However, one study detected the $^{675}$QTQTN$^{679}$-deleted mutants (mut-del2) in 3 of 68 SARS-CoV-2-positive clinical specimens (33). In another detection of 51 SARS-CoV-2-positive patient specimens, although a high rate of 52.9% and 82.4% of the positive clinical samples contained the FCS upstream motif ($^{661}$ECDIPIGAG$^{669}$) and the PRRA deletions, respectively, the mutant population was at a very low level (0.33% ±1.17% for FCS upstream motif deletion and 1.12% ±1.21% for PRRA deletion) (54), arguing for the infectivity and transmissibility of these mutants. Notably, we detected a high rate of mut-del2 (7.86% and 20.97%) but not mut-del1 (<0.1%) in two SARS-CoV-2-positive nasopharyngeal aspirates. Thus, it is possible that mut-del1 is an artificial deletion from the inoculated virus cultured in Vero-E6 cells. Therefore, further studies need to be carried out to understand the significance of the prevalence of these FCS deletion mutants in the disease progression of COVID-19 patients.

Along with the usages of antibody drugs and the wide inoculation of the vaccine, which target the S protein, the virus may undergo further mutations under the pressure of human immune response. Supervision and screening the mutations in the S protein gene in clinical specimens is extremely important to identify the escaped isolates which may increase or decrease infectivity and transmissibility. Apparently, the *in vitro*-polarized HAE model, which can facilitate long-term infection of SARS-CoV-2, is an ideal model to study S gene mutants under various conditions.

## MATERIALS AND METHODS

**Ethics statement.** Primary human bronchial epithelial cells were isolated from the lungs of healthy human donors and COPD patients by the Cells and Tissue Core of the Center for Gene Therapy, University of Iowa, and by the Department of Internal Medicine, University of Kansas Medical Center with the approvals of the institutional review boards (IRB) of the University of Iowa and University of Kansas Medical Center, respectively.

**Viruses.** SARS-CoV-2 (NR-52281), isolate USA-WA1/2020 (batch no. 70034262), was obtained from BEI Resources (Manassas, VA) and designated P0 passage. The virus used for infections of HAE-ALI was propagated once in Vero-E6 cells, designated P1 passage. Viruses were titrated by plaque assays on Vero-E6 cells and stored at −80°C as previously described (37). A biosafety protocol to work on SARS-CoV-2 infection in the biosafety level 3 (BSL3) lab was approved by the Institutional Biosafety Committee of the University of Kansas Medical Center.

**HAE-ALI cultures.** Primary HAE-ALI cultures, lots of B2-20, B3-20, B4-20, B9-20, B15-20, and B16-20, were directly prepared from bronchial airway epithelial cells isolated from various healthy donors. They were obtained from the Cells and Tissue Core of the Center for Gene Therapy, University of Iowa and polarized in Transwell inserts (0.33 $cm^2$; Costar, Corning, Tewksbury, WA). L209, KC19, and four COPD phenotypes (N707, UC24, L184, and N703) HAE-ALI cultures were prepared from propagated (passage 2) bronchial airway cells of the two healthy L209 and KC19 donors and four COPD donors. These six lung tissues were obtained from organ donors whose lungs were rejected for transplant and recovered for research by the Life Alliance Organ Recovery Agency at the University of Miami (Miami, Florida), Life Center Northwest (Seattle, Washington), and the Midwest Transplant Network (Kansas City, Kansas). Bronchial airway epithelia cells were polarized on Transwell inserts (1.1 $cm^2$) (Costar; Corning). The HAE-ALI cultures that had transepithelial electrical resistance (TEER) of >1,000 $\Omega \cdot cm^2$, determined with an epithelial volt-ohm meter (MilliporeSigma, Burlington, MA), were used for infections.

**Virus infections.** Polarized HAE-ALI cultures were infected with SARS-CoV-2 at a multiplicity of infection (MOI) of 0.2 or 2. The inoculum of 100 $\mu$l or 300 $\mu$l was apically applied to the 0.33-$cm^2$ or 1.1-$cm^2$ Transwell inserts with an incubation period of 1 h at 37°C and 5% $CO_2$. After aspiration of the inoculum, the apical surface of the insert was washed with 100 $\mu$l (or 300 $\mu$l) of Dulbecco's phosphate-buffered saline (D-PBS; Corning, Tewksbury, WA) three times to maximally remove the unbound viruses. The HAE-ALI cultures were then placed back into the incubator at 37°C and 5% $CO_2$. To collect the apically released progeny from infected cultures, 100 $\mu$l (or 300 $\mu$l) of D-PBS was added to the apical chamber for 30 min at 37°C and 5% $CO_2$. Thereafter, the apical wash was pipetted carefully from the apical chamber.

**Immunofluorescence assay.** The membrane of the infected HAE-ALI was cut out and fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) at 4°C overnight. The fixed membrane was washed in PBS for 5 min three times and then split into several pieces for whole-mount immunostaining. Following permeabilization with 0.2% Triton X-100 for 15 min at room temperature, the slide was incubated with a rabbit monoclonal anti-SARS-CoV-2 nucleocapsid (NP) (catalog no. 40143-R001; SinoBiological US, Wayne, PA) at a dilution of 1:25 in PBS with 2% fetal bovine serum for 1 h at 37°C. After washing, the slide was incubated with a rhodamine-conjugated secondary antibody, followed by staining of the nuclei with DAPI (4′,6-diamidino-2-phenylindole).

**RNA extraction.** For total RNA extraction, four Transwell inserts of HAE-ALI cultures were dissolved in 1 ml of TRIzol reagent (ThermoFisher, Waltham, MA), following manufacturer's instructions. Viral RNA

was isolated from the virions in apical washes. Fifty microliters of apical wash was used for the extraction of nuclease digestion-resistant viral RNA using the Quick-RNA Viral kit (catalog no. R1035; Zymo Research, Irvine, CA), as described previously (37). The final RNA samples were dissolved in 50 $\mu$l of deionized $H_2O$ and quantified for concentrations using a microplate reader (Synergy H; BioTek).

**RNA-seq.** For viral transcriptome, total RNA was extracted from HAE-ALI cultures infected with SARS-CoV-2 at an MOI of 0.2 and 2, respectively, or mock infected at 4 dpi. After RNA quality control and reverse transcription, DNA nanoball sequencing (DNSeq) was performed at BGI Genomics (Cambridge, MA). Briefly, RNA samples were tested using an Agilent 2100 bioanalyzer (Agilent RNA 6000 Nano kit). Samples with an RNA integrity number (RIN) of $\geq$8.0 were chosen for library construction. rRNA was removed from the total RNA samples by using RNase H or Ribo-Zero method. Then, samples were fragmented in a fragmentation buffer for thermal fragmentation to 130 to 160 nucleotides (nt). First-strand cDNA was generated by First Strand Mix, then Second Strand Mix was added to synthesize the second-strand cDNA. The reaction product was purified by magnetic beads and end repaired by addition of adaptors, followed by several rounds of PCR amplification to enrich the cDNA fragments. The PCR products were then purified and subjected to library quality control on the Agilent Technologies 2100 bioanalyzer. The double-stranded PCR products were heat denatured and circularized by the splint oligonucleotide sequence. The single-strand circle DNA (ssCir DNA) was formatted as the final library. The final library was amplified with phi29 to make DNA nanoballs (DNBs), which have more than 300 copies of one molecule. The DNBs were loaded into the patterned nanoarray and 2 $\times$ 100 paired-end reads were generated in the way of combinatorial probe-anchor synthesis (cPAS).

For RNA-seq of the viral RNAs, the apical washes were collected from infected HAE-ALI cultures at the indicated times (days postinfection [dpi]; Table 3), and viral RNA was extracted as described above. For library preparation, the stranded-RNA seq kit (Thermo Fisher) was used following the manufacturer's protocol. The rRNA depletion step was added for the library preparation. The Illumina sequencer NextSeq550 was used to generate paired-end 2 $\times$ 150 reads at GeneGoCell Inc. (San Diego, CA).

**PCR amplicon-seq.** For sequencing the FCS region of the S gene, viral RNA extracted from the apical washes was reverse transcribed using avian myeloblastosis virus (AMV) (Promega, Madison, WI). A 384-nt sequence covering the S gene FCS region (nt 23,487 to 23,870) was amplified by 20 cycles of PCR using the primers containing the adaptor sequences: forward, 5′-ACA CTC TTT CCC TAC ACG CTC TTC CGA TCT TTT TCA AAC ACG TGC AGG C-3′, and reverse, 5′-GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTT CCA GTT AAA GCA CGG TTT AAT-3′. The PCR products were analyzed on 1.5% agarose and excised for purification. The purified DNA samples were quantified on a microplate reader (Synergy LX; BioTek, Winooski, VT), and 500 ng of each DNA sample (20 ng/$\mu$l) was sent for PCR amplicon-seq (AMPLICON-EZ) at GENEWIZ, Inc. (South Plainfield, NJ).

**Bioinformatic analyses. (i) Total cellular DNBseq data (BGI) and PCR-amplicon-seq (GENEWIZ).** The reads were aligned to the reference SARS-CoV-2 Wuhan-Hu-1 isolate genome (GenBank accession no. MN908947) using BWA v0.7.5a-r405. Sequencing read coverage was calculated using bedtools genomecov of version 2.27.1. We used STAR (2.7.3a) to identify the junction-spanning reads as described previously except that we set the minimal size of deletions at 10 (15).

**(ii) Viral RNA-seq data (GeneGoCell Inc.).** Raw sequence reads (fastq files) were processed through the following steps by the Genenius NGS bioinformatics pipeline (v2.1). Low-quality reads were removed using a quality score threshold of 25 (Q25). The resulting fastq files were analyzed by FastQC v0.10.1 for quality control (QC). Reads were aligned to the reference genome Wuhan-Hu-1. The alignment results were analyzed using the proprietary GeneGoCell program for variant calling on the target sites as follows. (i) Each read pair was processed to report the variant in the read. (ii) Each variant's allele frequency (AF) was calculated based on the number of variant reads/total reads covering the region (both variant and nonvariant). (iii) Variants with $\geq$1% AF and $\geq$3 variant reads were reported in a variant calling file (vcf). The output of the bioinformatics workflow was collected and further organized/processed in Microsoft Office 365.

**Data availability.** All the RNA-seq and PCR amplicon-seq data have been deposited in NIH-sponsored BioProject database under accession number PRJNA698337 (https://dataview.ncbi.nlm.nih.gov/object/PRJNA698337?reviewer=k4gtr6eundj03jnpcrj62tq1un).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**DATA SET S1**, XLSX file, 0.9 MB.

## ACKNOWLEDGMENTS

We thank the Cells and Tissue Core of Center for Gene Therapy, the University of Iowa (DK054759) for providing the primary cell cultures and thank Rachael Liesman for providing SARS-CoV-2-positive nasopharyngeal specimens. The following reagent was deposited by the Centers for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH: SARS-related coronavirus 2, isolate USA-WA1/2020, NR-52281.

The study was supported by PHS grant AI150877 and an internal award from the University of Kansas Medical Center.

Elizabeth Yan Zhang is the founder of GeneGoCell Inc.

## REFERENCES

1. Hui DS, Azhar I, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, Mchugh TD, Memish ZA, Drosten C, Zumla A, Petersen E. 2020. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - the latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis 91:264–266. https://doi.org/10.1016/j.ijid.2020.01.009.
2. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273. https://doi.org/10.1038/s41586-020-2012-7.
3. 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. 2020. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382:727–733. https://doi.org/10.1056/NEJMoa2001017.
4. Krishnan A, Hamilton JP, Alqahtani SA, Woreta TA. 2021. COVID-19: an overview and a clinical update. World J Clin Cases 9:8–23. https://doi.org/10.12998/wjcc.v9.i1.8.
5. Lamb YN. 2020. Remdesivir: first approval. Drugs 80:1355–1363. https://doi.org/10.1007/s40265-020-01378-w.
6. Chung JY, Thone MN, Kwon YJ. 2021. COVID-19 vaccines: the status and perspectives in delivery points of view. Adv Drug Deliv Rev 170:1–25. https://doi.org/10.1016/j.addr.2020.12.011.
7. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5:536–544. https://doi.org/10.1038/s41564-020-0695-z.
8. Wu Y, Ho W, Huang Y, Jin DY, Li S, Liu SL, Liu X, Qiu J, Sang Y, Wang Q, Yuen KY, Zheng ZM. 2020. SARS-CoV-2 is an appropriate name for the new coronavirus. Lancet 395:949–950. https://doi.org/10.1016/S0140-6736(20)30557-2.
9. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395:565–574. https://doi.org/10.1016/S0140-6736(20)30251-8.
10. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 9:221–236. https://doi.org/10.1080/22221751.2020.1719902.
11. Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, Ahuja A, Yung MY, Leung CB, To KF, Lui SF, Szeto CC, Chung S, Sung JJ. 2003. A major outbreak of severe acute respiratory syndrome in Hong Kong. N Engl J Med 348:1986–1994. https://doi.org/10.1056/NEJMoa030685.
12. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Berger A, Burguiere AM, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra JC, Muller S, Rickerts V, Sturmer M, Vieth S, Klenk HD, Osterhaus AD, Schmitz H, Doerr HW. 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 348:1967–1976. https://doi.org/10.1056/NEJMoa030747.
13. Miao Z, Tidu A, Eriani G, Martin F. 2020. Secondary structure of the SARS-CoV-2 5'-UTR. RNA Biol 18:447–456. https://doi.org/10.1080/15476286.2020.1814556.
14. Zhao J, Qiu J, Aryal S, Hackett JL, Wang J. 2020. The RNA architecture of the SARS-CoV-2 3'-untranslated region. Viruses 12:1473. https://doi.org/10.3390/v12121473.
15. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. 2020. The architecture of SARS-CoV-2 transcriptome. Cell 181:914–921. https://doi.org/10.1016/j.cell.2020.04.011.
16. Irigoyen N, Firth AE, Jones JD, Chung BY, Siddell SG, Brierley I. 2016. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. PLoS Pathog 12:e1005473. https://doi.org/10.1371/journal.ppat.1005473.
17. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom KJ, Zambon M, Ellis J, Lewis PA, Hiscox JA, Matthews DA. 2020. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. Genome Med 12:68–763. https://doi.org/10.1186/s13073-020-00763-0.
18. Sawicki SG, Sawicki DL, Siddell SG. 2007. A contemporary view of coronavirus transcription. J Virol 81:20–29. https://doi.org/10.1128/JVI.01358-06.
19. Sola I, Almazán F, Zúñiga S, Enjuanes L. 2015. Continuous and discontinuous RNA synthesis in coronaviruses. Annu Rev Virol 2:265–288. https://doi.org/10.1146/annurev-virology-100114-055218.
20. Zúñiga S, Sola I, Alonso S, Enjuanes L. 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. J Virol 78:980–994. https://doi.org/10.1128/JVI.78.2.980-994.2004.
21. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. 2020. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol 19:155−170. https://doi.org/10.1038/s41579-020-00468-6.
22. Fehr AR, Perlman S. 2015. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol 1282:1–23. https://doi.org/10.1007/978-1-4939-2438-7_1.
23. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL, Tummino TA, Hüttenhain R, Kaake RM, Richards AL, Tutuncuoglu B, Foussard H, Batra J, Haas K, Modak M, Kim M, Haas P, Polacco BJ, Braberg H, Fabius JM, Eckhardt M, Soucheray M, Bennett MJ, Cakir M, McGregor MJ, Li Q, Meyer B, Roesch F, Vallet T, Mac Kain A, Miorin L, Moreno E, Naing ZZC, Zhou Y, Peng S, Shi Y, Zhang Z, Shen W, Kirby IT, Melnyk JE, Chorba JS, Lou K, Dai SA, Barrio-Hernandez I, Memon D, Hernandez-Armenta C, et al. 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583:459–468. https://doi.org/10.1038/s41586-020-2286-9.
24. Li F. 2016. Structure, function, and evolution of coronavirus spike proteins. Annu Rev Virol 3:237–261. https://doi.org/10.1146/annurev-virology-110615-042301.
25. Hu B, Guo H, Zhou P, Shi ZL. 2020. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol 19:141−154. https://doi.org/10.1038/s41579-020-00459-7.
26. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, Li F. 2020. Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci U S A 117:11727–11734. https://doi.org/10.1073/pnas.2003138117.
27. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann 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.
28. Hoffmann M, Kleine-Weber H, Pöhlmann 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.
29. Zhou H, Chen X, Hu T, Li J, Song H, Liu Y, Wang P, Liu D, Yang J, Holmes EC, Hughes AC, Bi Y, Shi W. 2020. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. Curr Biol 30:2196–2203. https://doi.org/10.1016/j.cub.2020.05.023.
30. Jaimes JA, Millet JK, Whittaker GR. 2020. Proteolytic cleavage of the SARS-CoV-2 spike protein and the role of the novel S1/S2 site. iScience 23:101212. https://doi.org/10.1016/j.isci.2020.101212.
31. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. 2020. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res 176:104742. https://doi.org/10.1016/j.antiviral.2020.104742.
32. Klimstra WB, Tilston-Lunel NL, Nambulli S, Boslett J, McMillen CM, Gilliland T, Dunn MD, Sun C, Wheeler SE, Wells A, Hartman AL, McElroy AK, Reed DS, Rennick LJ, Duprex WP. 2020. SARS-CoV-2 growth, furin-cleavage-site adaptation and neutralization using serum from acutely

infected hospitalized COVID-19 patients. J Gen Virol 101:1156–1169. https://doi.org/10.1099/jgv.0.001481.
33. Liu Z, Zheng H, Lin H, Li M, Yuan R, Peng J, Xiong Q, Sun J, Li B, Wu J, Yi L, Peng X, Zhang H, Zhang W, Hulswit RJG, Loman N, Rambaut A, Ke C, Bowden TA, Pybus OG, Lu J. 2020. Identification of common deletions in the spike protein of severe acute respiratory syndrome coronavirus 2. J Virol 94:e00790-20. https://doi.org/10.1128/JVI.00790-20.
34. 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.
35. Ogando NS, Dalebout TJ, Zevenhoven-Dobbe JC, Limpens RWAL, van der Meer Y, Caly L, Druce J, de Vries JJC, Kikkert M, Bárcena M, Idorov I, Nijder EJ. 2020. SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. J Gen Virol 101:925–940. https://doi.org/10.1099/jgv.0.001453.
36. Zhu N, Wang W, Liu Z, Liang C, Wang W, Ye F, Huang B, Zhao L, Wang H, Zhou W, Deng Y, Mao L, Su C, Qiang G, Jiang T, Zhao J, Wu G, Song J, Tan W. 2020. Morphogenesis and cytopathic effect of SARS-CoV-2 infection in human airway epithelial cells. Nat Commun 11:3910. https://doi.org/10.1038/s41467-020-17796-z.
37. Hao S, Ning K, Kuz CA, Vorhies K, Yan Z, Qiu J. 2020. Long-term modeling of SARS-CoV-2 infection of in vitro cultured polarized human airway epithelium. mBio 11:e02852-20. https://doi.org/10.1128/mBio.02852-20.
38. Sethna PB, Hofmann MA, Brian DA. 1991. Minus-strand copies of replicating coronavirus mRNAs contain antileaders. J Virol 65:320–325. https://doi.org/10.1128/JVI.65.1.320-325.1991.
39. Harilal D, Ramaswamy S, Loney T, Suwaidi HA, Khansaheb H, Alkhaja A, Varghese R, Deesi Z, Nowotny N, Alsheikh-Ali A, Abou TA. 2020. SARS-CoV-2 whole genome amplification and sequencing for effective population-based surveillance and control of viral transmission. Clin Chem 66:1450–1458. https://doi.org/10.1093/clinchem/hvaa187.
40. Butler DJ, Mozsary C, Meydan C, Danko D, Foox J, Rosiene J, Shaiber A, Afshinnekoo E, MacKay M, Sedlazeck FJ, Ivanov NA, Sierra M, Pohle D, Zietz M, Gisladottir U, Ramlall V, Westover CD, Ryon K, Young B, Bhattacharya C, Ruggiero P, Langhorst BW, Tanner N, Gawrys J, Meleshko D, Xu D, Steel PAD, Shemesh AJ, Xiang J, Thierry-Mieg J, Thierry-Mieg D, Schwartz RE, Iftner A, Bezdan D, Sipley J, Cong L, Craney A, Velu P, Melnick AM, Hajirasouliha I, Horner SM, Iftner T, Salvatore M, Loda M, Westblade LF, Cushing M, Levy S, Wu S, Tatonetti N, Imielinski M, et al. 2020. Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. Nat Commun 12:1660. https://doi.org/10.1038/s41467-021-21361-7.
41. Cantuti-Castelvetri L, Ojha R, Pedro LD, Djannatian M, Franz J, Kuivanen S, van der Meer F, Kallio K, Kaya T, Anastasina M, Smura T, Levanov L, Szirovicza L, Tobi A, Kallio-Kokko H, Österlund P, Joensuu M, Meunier FA, Butcher SJ, Winkler MS, Mollenhauer B, Helenius A, Gokce O, Teesalu T, Hepojoki J, Vapalahti O, Stadelmann C, Balistreri G, Simons M. 2020. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. Science 370:856–860. https://doi.org/10.1126/science.abd2985.
42. Sasaki M, Uemura K, Sato A, Toba S, Sanaki T, Maenaka K, Hall WW, Orba Y, Sawa H. 2021. SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated in vitro during propagation in TMPRSS2-deficient cells. PLoS Pathog 17:e1009233. https://doi.org/10.1371/journal.ppat.1009233.
43. Tang T, Jaimes JA, Bidon MK, Straus MR, Daniel S, Whittaker GR. 2021. Proteolytic activation of SARS-CoV-2 spike at the S1/S2 boundary: potential role of proteases beyond furin. ACS Infect Dis 7:264–272. https://doi.org/10.1021/acsinfecdis.0c00701.
44. Zhu Y, Feng F, Hu G, Wang Y, Yu Y, Zhu Y, Xu W, Cai W, Sun Z, Han W, Ye R, Chen H, Ding Q, Cai Q, Qu D, Xie Y, Yuan Z, Zhang R. 2021. A genome-wide screen identifies host factors that regulate SARS-CoV-2 entry. Nat Commun 12:961. https://doi.org/10.1038/s41467-021-21213-4.
45. Ren X, Glende J, Al-Falah M, de Vries V, Schwegmann-Wessels C, Qu X, Tan L, Tschernig T, Deng H, Naim HY, Herrler G. 2006. Analysis of ACE2 in polarized epithelial cells: surface expression and function as receptor for severe acute respiratory syndrome-associated coronavirus. J Gen Virol 87:1691–1695. https://doi.org/10.1099/vir.0.81749-0.
46. Böttcher-Friebertshäuse E, Stein DA, Klenk HD, Garten W. 2011. Inhibition of influenza virus infection in human airway cell cultures by an antisense peptide-conjugated morpholino oligomer targeting the hemagglutinin-activating protease TMPRSS2. J Virol 85:1554–1562. https://doi.org/10.1128/JVI.01294-10.
47. Ziegler CGK, Allon SJ, Nyquist SK, Mbano IM, Miao VN, Tzouanas CN, Cao Y, Yousif AS, Bals J, Hauser BM, Feldman J, Muus C, Wadsworth MH, Kazer SW, Hughes TK, Doran B, Gatter GJ, Vukovic M, Taliaferro F, Mead BE, Guo Z, Wang JP, Gras D, Plaisant M, Ansari M, Angelidis I, Adler H, Sucre JMS, Taylor CJ, Lin B, Waghray A, Mitsialis V, Dwyer DF, Buchheit KM, Boyce JA, Barrett NA, Laidlaw TM, Carroll SL, Colonna L, Tkachev V, Peterson CW, Yu A, Zheng HB, Gideon HP, Winchell CG, Lin PL, Bingle CD, Snapper SB, Kropski JA, Theis FJ, et al. 2020. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. Cell 181:1016–1035. https://doi.org/10.1016/j.cell.2020.04.035.
48. Tang T, Bidon M, Jaimes JA, Whittaker GR, Daniel S. 2020. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. Antiviral Res 178:104792. https://doi.org/10.1016/j.antiviral.2020.104792.
49. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, Chen R, Decramer M, Fabbri LM, Frith P, Halpin DMG, López Varela MV, Nishimura M, Roche N, Rodriguez-Roisin R, Sin DD, Singh D, Stockley R, Vestbo J, Wedzicha JA, Agusti A. 2017. Global strategy for the diagnosis, management and prevention of chronic obstructive lung disease 2017 report: GOLD executive summary. Respirology 22:575–601. https://doi.org/10.1111/resp.13012.
50. Alberca RW, Lima JC, de Oliveira EA, Gozzi-Silva SC, Ramos YÃ, Andrade MMS, Beserra DR, Oliveira LM, Branco ACCC, Pietrobon AJ, Pereira NZ, Teixeira FME, Fernandes IG, Duarte AJDS, Benard G, Sato MN. 2021. COVID-19 disease course in former smokers, smokers and COPD patients. Front Physiol 11:637627. https://doi.org/10.3389/fphys.2020.637627.
51. Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, Deng Y, Lin S. 2020. The impact of COPD and smoking history on the severity of COVID-19: a systemic review and meta-analysis. J Med Virol 92:1915–1921. https://doi.org/10.1002/jmv.25889.
52. Johnson BA, Xie X, Bailey AL, Kalveram B, Lokugamage KG, Muruato A, Zou J, Zhang X, Juelich T, Smith JK, Zhang L, Bopp N, Schindewolf C, Vu M, Vanderheiden A, Winkler ES, Swetnam D, Plante JA, Aguilar P, Plante KS, Popov V, Lee B, Weaver SC, Suthar MS, Routh AL, Ren P, Ku Z, An Z, Debbink K, Diamond MS, Shi PY, Freiberg AN, Menachery VD. 2021. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. Nature 591:293−299. https://doi.org/10.1038/s41586-021-03237-4.
53. Wan P, Lau SY, Deng S, Chen P, Mok BW, Zhang AJ, Lee AC, Chan KH, Song W, To KK, Chan JF, Yuen KY, Chen H. 2020. Pathogenicity, immunogenicity, and protective ability of an attenuated SARS-CoV-2 variant with a deletion at the S1/S2 junction of the spike protein. bioRxiv https://doi.org/10.1101/2020.08.24.264192.
54. Wong YC, Lau SY, Wang To KK, Mok BWY, Li X, Wang P, Deng S, Woo KF, Du Z, Li C, Zhou J, Woo Chan JF, Yuen KY, Chen H, Chen Z. 2020. Natural transmission of bat-like SARS-CoV-2ΔPRRA variants in COVID-19 patients. Clin Infect Dis https://doi.org/10.1093/cid/ciaa953.