Nuclear translocation of spike mRNA and protein is a novel pathogenic feature of SARS-CoV-2.

Sarah Sattar[1#], Juraj Kabat[2#], Kailey Jerome[1], Friederike Feldmann[3], Kristina Bailey[4], and Masfique Mehedi[1#*]

[1]Department of Biomedical Sciences, University of North Dakota School of Medicine & Health Sciences, Grand Forks, ND, USA.

[2]Biological Imaging Section, Research Technology Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA.

[3]Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, USA.

[4]Department of Internal Medicine, Pulmonary, Critical Care, and Sleep and Allergy, University of Nebraska Medical Center, Omaha, NE, USA.

#Contributed equally

[*]Correspondence author

Email: masfique.mehedi@und.edu

Short title: S mRNA and S protein colocalize and translocate into the nucleus

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes severe pathophysiology in vulnerable older populations and appears to be highly pathogenic and more transmissible than SARS-CoV or MERS-CoV [1, 2]. The spike (S) protein appears to be a major pathogenic factor that contributes to the unique pathogenesis of SARS-CoV-2. Although the S protein is a surface transmembrane type 1 glycoprotein, it has been predicted to be translocated into the nucleus due to the novel nuclear localization signal (NLS) "PRRARSV", which is absent from the S protein of other coronaviruses. Indeed, S proteins translocate into the nucleus in SARS-CoV-2-infected cells. To our surprise, S mRNAs also translocate into the nucleus. S mRNA colocalizes with S protein, aiding the nuclear translocation of S mRNA. While nuclear translocation of nucleoprotein (N) has been shown in many coronaviruses, the nuclear translocation of both S mRNA and S protein reveals a novel pathogenic feature of SARS-CoV-2.

Author summary

One of the novel sequence insertions resides at the S1/S2 boundary of Spike (S) protein and constitutes a functional nuclear localization signal (NLS) motif "PRRARSV", which may supersede the importance of previously proposed polybasic furin cleavage site "RRAR". Indeed, S protein's NLS-driven nuclear translocation and its possible role in S mRNA's nuclear translocation reveal a novel pathogenic feature of SARS-CoV-2.

Introduction

The recently emerged severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), along with SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV), belong to the *Coronaviridae* virus family. The current ongoing outbreak has shown that SARS-CoV-2 is highly pathogenic and more transmissible than SARS-CoV or MERS-CoV [1]. These coronaviruses contain a positive-strand RNA genome with a few unique features: two-thirds of the viral RNA is translated into a large polyprotein, and the remainder of the viral genome is transcribed by a discontinuous transcription process into a nested set of subgenomic mRNAs [3-5]. The different subgenomic RNAs encode four conserved structural proteins (spike, S; envelope, E; membrane, M, and nucleocapsid, N) and several accessory proteins [6, 7]. The S protein of both SARS-CoV and SARS-CoV-2 interacts with the host cell receptor angiotensin converting enzyme 2 (ACE2) and triggers fusion between the viral envelope and host cell membrane to facilitate successful viral entry [8, 9]. However, the S protein of MERS-CoV binds to dipentidyl peptidase (DPP4) to facilitate entry into cells [10]. Importantly, the SARS-CoV-2 S protein is a significant pathogenic factor because of its broad tropism for mammalian ACE2 [11]. While the S protein is an attractive target for therapeutic development [12], the lack of comprehensive information on S protein expression and subcellular translocation hinders the identification of an effective S protein-targeting therapeutic to combat SARS-CoV-2 infection.

The genome sequence is generally the blueprint for detecting biological function [13]. Thus, the S protein's function is encoded in the S gene sequence. Identifying novel features in the S gene sequence, its expression and subcellular localization may shed light on the unique pathogenesis of SARS-CoV-2 compared to other pathogenic beta-coronaviruses, particularly SARS-CoV and

MERS-CoV. A recent study showed several SARS-CoV-2 genomic features, including novel sequence insertions and enhanced N protein nuclear localization signals (NLSs) that are thought to be responsible for the unique pathogenesis of this coronavirus [14]. There are three types of NLSs: pat4, pat7, and bipartite. The pat4 signal is a chain of 4 basic amino acids consisting of lysine or arginine or three basic amino acids, with the last amino acid being either histidine or proline. The pat7 signal begins with proline and is followed by six amino acids, which contains a four-residue sequence in which three of the four residues are basic. The bipartite signal consists of two basic amino acids with a 10-residue spacer and a five amino acid sequence in which at least three of the five amino acids are basic [15-17]. The subcellular localization of some SARS-CoV-2 proteins has been studied in vitro [18], but a comprehensive understanding of the subcellular localization of the S protein is missing.

Here, we first report the nuclear translocation of S protein and mRNA in SARS-CoV-2-infected cells. The translocation of the SARS-CoV-2 S mRNA appeared to be assisted by the S protein, which contains an NLS motif that is unique among human pathogenic beta-coronaviruses.

Results

The novel NLS motif "PRRARSV" is in the S protein of SARS-CoV-2 but not SARS-CoV or MERS-CoV.

Several groups have reported novel nucleotide insertions in the S gene of SARS-CoV-2, as indicated by a multiple sequence alignment for the S protein sequences of different coronaviruses, such as a polybasic site "PRRA" produced by a 12-nucleotide acquisition at the S1-S2 boundary

through multiple host-species adaptations [19, 20]. However, S protein sequence alignments between SARS-CoV-2 and SARS-CoV showed the possibility of the insertions "NSPR" [21] and "SPRR" [22] at the S1-S2 boundary. It has previously been reported that the sequence insertion at the S1-S2 boundary constitutes a furin cleavage site [23, 24]. A comprehensive understanding of the consequence of the sequence insertion at the S1-S2 boundary is still missing, possibly because research focused on understanding the differences in the pathogenicity of the different SARS-CoV-2 variants and subvariants, which emerged rapidly. To determine whether the earlier SARS-CoV-2 isolate (USA/WA-CDC-WA1/2020 isolate, GenBank accession no. MN985325) has multiple novel sequence insertions in the S protein compared to SARS-CoV (Urbani strain, GenBank accession no. AY278741), we aligned the S protein sequences of both viruses using a constraint-based alignment tool for multiple protein sequences (COBALT) [25]. We did not use MER-CoV for comparison because there is only 40% similarity between SARS-CoV-2 and MERS-CoV [26]. Similar to a previous report [21], we found sequence insertions (IS) in the SARS-CoV-2 S protein at four independent positions: IS1 "GTNGKTR", IS2 "YYHK", IS3 "HRSY", and IS4 "NSPR" (Fig. 1A & B). To determine whether any of these sequence insertions constituted or resembled any protein motifs, such as an NLS, we analyzed the SARS-CoV-2 S protein in silico with the PSORT II web portal for NLS prediction [27]. We found that the SARS-CoV-2 glycoprotein contained an NLS of the "pat7" motifs, one of the three NLS motifs described above (Fig. S1). To our surprise, the NLS motif "PRRARSV" was present at the proposed polybasic site and was due to the fourth sequence insertion "NSPR" (Figs. 1A & B and S1). A widely reported furin consensus cleavage site motif is the canonical four amino acid motif R-X-[K/R]-R, although R-X-X-R is the minimal cleavage site on the substrate for successful furin cleavage [28, 29]. Due to the specificity of the amino acid motif, a furin cleavage motif is not

expected to fulfill the characteristics of an NLS motif. However, the described furin cleavage site is constitutively within the NLS motif. Thus, whether furin cleavage destroys the function of the NLS motif is important to determine. As expected, the NLS in the S protein was unique to SARS-CoV-2 among human pathogenic beta-coronaviruses, as neither the SARS-CoV S protein nor the MERS-CoV S protein has an NLS (Fig. S1).

NLS-driven nuclear translocation of S protein (including S mRNA) occurs only in the SARS-CoV-2-infected airway epithelium.

Although viral glycoprotein nuclear translocation is rare, NLS-driven protein nuclear translocation has already been established in different viral infections [30, 31]. Thus, it is important to determine whether the SARS-CoV-2 S protein translocates into the nucleus in addition to its canonical cell surface localization through the ER-Golgi pathway. We hypothesized that the S protein could translocate into the nucleus in SARS-CoV-2-infected cells via the identified NLS motif [17, 30, 31]. We infected highly differentiated pseudostratified airway epithelial cells (which mimics *in vivo* human airway epithelium) with SARS-CoV-2 at a multiplicity of infection (MOI) of 0.1 for four days. First, we confirmed the presence of S mRNA and S protein in a 5 µm section of formalin-fixed paraffin-embedded SARS-CoV-2-infected cells by RNAscope and immunofluorescence analysis. Despite the rarity of viral mRNA (or even positive-strand RNA virus genome) to be nuclear [32, 33], a recent study showed that SARS-CoV-2 mRNA accumulates in the nucleus of infected cells [34]. Our results showed that SARS-CoV-2 S mRNA was nuclear (Fig. 2 left panel and merged images in the right panel). To confirm the physical apposition between S mRNA and the nucleus by comparing their distributions in fluorescent images, we used the spot-to-spot colocalization function in Imaris image analysis software (Oxford Instruments).

We found that S mRNA was nuclear and abundant in the cytoplasm (Fig. S2, top panel, three donors). To avoid image artifacts, we imaged multiple independent slides of SARS-CoV-2-infected airway epithelium (from three independent donors) using at least two different high-end confocal microscopes. Additionally, we used at least two different image processing strategies to determine nuclear localization. Based on high-resolution imaging, we determined the subcellular distribution of S mRNA at the single-molecule and single-cell levels (Figs. 2 & 3A, S2 & S3). Importantly, we were able to determine S mRNA nuclear translocation not only inside the nucleus but also on the nuclear surface (Figs. 3A, 3C, and S3). The determination of S mRNA distribution and abundance showed that S mRNA subcellular localization spans from the inside and outer surface of the nuclear membrane to everywhere in the cytoplasm. We found that almost 90% of S mRNA was distributed in the cytoplasm, which was expected, as SARS-CoV-2 transcription and replication occur in the cytoplasm (Figs. 3 A & C, S3). Interestingly, less than 10% of S mRNA was detected at the nuclear surface, which could explain the transitionary stage of S mRNA before it enters the nucleus or the novel transnuclear-membrane translocation of S mRNA, which was examined and described later (Figs. 3 A & C, S3). In approximately 1% of instances, S mRNA successfully translocated into the nucleus (Figs. 3 A & C, S3). The nuclear translocation of S mRNA is highly unusual because there have been few previous reports of S mRNA nuclear translocation and no information on the mechanism of nuclear translocation. However, we have explored how S mRNA could translocate into the nucleus.

We investigated whether the S protein translocated into the nucleus in the SARS-CoV-2-infected airway epithelium. Consistent with the S mRNA data, we found that the S protein translocated into the nucleus and was abundant on the cellular surface through the cytoplasmic ER-Golgi pathway

(Fig. 2, middle panel and merged images in the right panel). Based on high-resolution imaging, we determined that S protein nuclear translocation included both the inside of the nucleus and the nuclear surface (Figs. 3B, 3C, and S3). Similar to S mRNA quantification, we were able to quantify the distribution and abundance of S protein in the infected airway epithelium. We found that the S protein was distributed inside and in the outer membrane of the nucleus, the cytoplasmic ER-Golgi and the cell surface. We did not determine what portion of the S protein was localized inside and outside the cell surface. However, we quantified the subcellular distribution of the S protein inside the nucleus, outside the surface of the nucleus, and in the cytoplasm, which included cell surface expression because the S protein is a type 1 transmembrane glycoprotein. We found that approximately 75% of the S protein was distributed in sites other than the nucleus, including the cell surface and the cytoplasm, which was expected, as S protein translation and protein processing occur in the cytoplasm and via cytoplasmic ER-Golgi pathway, respectively (Figs. 3 B & C, S3). Interestingly, approximately 15% of the S protein was detected at the nuclear surface, which could explain the S protein transitionary stage before entering the nucleus or a novel transnuclear-membrane translocation of S protein, which was examined and described later (Figs. 3 A & C, S3). Interestingly, we found that a higher percentage of total S protein translocated into the nucleus than S mRNA (Figs. 3 A-C, S3). Although viral type-1 transmembrane glycoprotein translocation into the nucleus is rare, the NLS in the S protein is responsible for nuclear translocation. It was apparent that NLS-driven S protein nuclear translocation was SARS-CoV-2 specific, and a side-by-side infection experiment with both viruses showed that the S protein of SARS-CoV did not translocate into the nucleus (Fig. S4). As both S mRNA and S protein translocated into the nucleus, it is important to determine whether S mRNA and S protein colocalize in different subcellular sites.

176

177 **Colocalization of S mRNA and S protein in different subcellular locations in the SARS-CoV-**
178 **2-infected airway epithelium.**

179 While we can explain S protein nuclear translocation due to the presence of an NLS motif in the
180 amino acid sequence, we can only hypothesize that S mRNA nuclear translocation is possible due
181 to a direct interaction between S protein and S mRNA, which can be explained by the
182 colocalization between them. The SARS-CoV-2 N protein is an abundant RNA binding protein
183 that is essential for viral genome packaging [35]. While the structural basis of N protein binding
184 to single- or double-stranded RNA is known [36], there is no information about whether S protein
185 binds to S mRNA. As we found similar intracellular distribution of both S mRNA and S protein,
186 we hypothesized that S protein interacts with S mRNA to translocate the protein–mRNA complex
187 to different subcellular locations, including the cytoplasm and nucleus, but not the cell surface. By
188 examining the colocalization between the S protein and S mRNA, we could confirm the presence
189 of the protein–mRNA complex in the SARS-CoV-2-infected airway epithelium. Here, we refer to
190 colocalization as an association between S mRNA and S protein at different intracellular locations.
191 Technically, two separate fluorescence molecules that emit different wavelengths of light are
192 superimposed within an indeterminate microscope resolution. To determine whether S mRNA and
193 S protein colocalize, we used a high-resolution imaging strategy. We quantified the colocalization
194 on a percentage scale. We found that approximately 85% of the colocalization, which was the
195 highest, was observed outside the nucleus (Fig. 3D). These data are consistent with the previously
196 described spatial expression data of both S protein and S mRNA (Fig. 3A-C). As expected, lower
197 and the lowest percentages of colocalization between the S protein and S mRNA were observed
198 on the nuclear surface and inside the nucleus, respectively (Fig. 3D). We were able to pinpoint the

199 colocalization site by using high magnification confocal imaging followed by image processing.
200 Representatives of S mRNA and S protein colocalization in the cytoplasm (Fig 4, top two panels),
201 on the surface of the nuclear membrane (Fig 4, middle two panels), or inside the nucleus (Fig 4,
202 bottom two panels) are shown. We observed S protein and S mRNA colocalization in three
203 subcellular locations, which was confirmed at the single-cell level in SARS-CoV-2-infected cells
204 (Fig. S5). We also observed that S mRNA inside and on the nuclear surface was associated with
205 the S protein, in contrast to the cytoplasmic S mRNA distribution with or without colocalization
206 with the S protein (Figs. 4 and S5). Thus, S mRNA translocates into the nucleus through the S
207 protein-S mRNA complex and is driven by the S protein (Figs. 4 and S5).

208

209 **NLS-driven N protein nuclear translocation is common in SARS-CoV, MERS-CoV, and**
210 **SARS-CoV-2 infections.**

211 There is no comprehensive information on SARS-CoV-2 N protein NLS motifs, and we already
212 know that other pathogenic coronaviruses, particularly SARS-CoV [17, 37] and MERS-CoV [38]
213 N proteins, have NLSs. Therefore, we searched for NLSs in the SARS-CoV-2 N protein by using
214 the PSORT II. We found that the SARS-CoV-2 N protein has 7 NLSs covering all three types of
215 NLS motifs (pat4: 2; pat7: 3, and bipartite: 2); however, the SARS-CoV N protein has 8 NLS
216 motifs (pat4: 2; pat7: 4, and bipartite: 2) (Fig. S6). N protein translocation into the nucleus or at
217 least the perinuclear region was confirmed (Figs. 5A and S7). However, SARS-CoV-2 N protein
218 nuclear translocation was not as robust as SARS-CoV nuclear translocation (Figs. 5B and S7). The
219 reduced nuclear translocation of the SARS-CoV-2 N protein is probably due to the absence of one
220 pat7 NLS motif in the SARS-CoV-2 N protein compared to that in SARS-CoV (Figs. S7 and S8).

221

222 **SARS-CoV-2 S and N proteins' interactions with genomic RNA.**

223 Based on a machine learning model, the SARS-CoV-2 RNA genome and sub-genomic RNAs can
224 be translocated in the host cells' mitochondrial matrix and nucleus [39]. Our results suggest that
225 around 1% S mRNA translocated into the nucleus. S mRNA's (potentially SARS-CoV-2 mRNA
226 genome) subcellular localization maqy play a significant role in SARS-CoV-2 pathogenesis. To
227 determine whether the SARS-CoV-2 genome interacts with either S protein or N protein, we in
228 silico analyzed RNA-protein interactions using the RPISeq web portal, which offers the only
229 sequence-based prediction model [40]. We found that both S and N protein binding probability to
230 the SARS-CoV-2 genome scored exactly 1 (Dataset S1 and S2). SARS-CoV-2 N protein is an
231 abundant RNA binding protein essential for viral genome packaging [35]. While the structural
232 basis of N binding to single or double-stranded RNA is known [36], we found that S mRNA
233 nuclear translocation aids via S protein. However, the mechanism of S protein binding to the
234 mRNA or possibly positive-strand RNA genome is yet to be determined.

235

236 **Discussion**

237 In the context of SARS-CoV, one of the controversies regarding the natural origin of SARS-CoV-2
238 is that its S gene has multiple novel sequence insertions. Zhang C. et al. analyzed the report by
239 Pradhan et al. (withdrawn) [41] on the presence of four unique novel sequences in the SARS-CoV-
240 2 S gene and showed that these four sequence insertions were not related to the receptor-binding
241 domain (RBD) [21]. A recent study identified S gene novel sequence insertions among several key
242 genomic features that differentiate SARS-CoV-2 from other beta-coronaviruses, particularly

243 SARS-CoV and MERS-CoV [14]. The source and characterization of these sequence insertions
244 have yet to be determined; however, the closest BLAST hit of these sequences is bat coronavirus
245 RaTG13 [42]. Similar to a previous report [21], we found four multiple sequence insertions in the
246 SARS-CoV-2 S protein: IS1 "GTNGKTR", IS2 "YYHK", "HRSY", and IS4 "NSPR" (Fig. 1).
247 Here, we showed that the fourth novel sequence insertion in the S gene was an NLS and resulted
248 in nuclear translocation of the S protein, which not only complemented previous in silico findings
249 [14] but also identified a novel pathogenic genomic feature of the S gene. Interestingly, the fourth
250 significant insertion has received attention due to the description of a polybasic site "RRAR",
251 which may contribute to increased serin protease-driven entry of SARS-CoV-2 [19] and is
252 implicated in broader tropism and/or enhanced viral transmissibility compared to SARS-CoV [20].
253 However, we found that the IS4 "NSPR" created a pat7 NLS "PRRARSV" in the S protein, which
254 was unique to SARS-CoV-2. We first reported that the S protein translocated into the nucleus in
255 the SARS-CoV-2-infected airway epithelium, which is an appropriate lung model for studying
256 respiratory virus infection in vitro [2, 43]. Our results confirmed that the SARS-CoV-2 S protein
257 was a unique addition to the list of viral proteins that possess NLSs and consequently translocate
258 into the nucleus of infected cells [17, 30, 31, 44]. Among coronaviruses, SARS-CoV-2 S protein
259 is the first type-1 transmembrane glycoprotein that translocates into the nucleus. Vesicular
260 stomatitis virus (VSV), which is a negative sense RNA virus, has a glycoprotein that translocates
261 to the nucleus as well. A study by the University of Illinois at Urbana Champaign showed exactly
262 how the glycoprotein on VSV was able to travel to the nucleus of hamster kidney cells [45].

263 NLS-driven S protein nuclear translocation is a novel pathogenic feature of SARS-CoV-2
264 infection compared to other pathogenic coronaviruses. However, the pathogenic contribution of
265 the S protein's NLS motif to virus-induced pathophysiology is yet to be determined. Our results

suggested that the S protein translocated into the nucleus due to the NLS, which also raised two important points. First, we investigated whether the proposed polybasic site "RRAR" could itself be an NLS motif. The answer was that the proposed polybasic site was not an NLS motif because an NLS is a well characterized and predefined amino acid sequence motif [15-17]. Additionally, the amino acid sequence of the probable sequence insertion "NSPR" [21] was also not an NLS but was part of the P7 "PRRARSV" NLS. Thus, the inserted sequence creates the NLS in the S protein of SARS-CoV-2 and may make SARS-CoV-2 unique among human pathogenic coronaviruses.

The second important point was whether the NLS motif was functional in the context of the described polybasic site at the S1/S2 boundary. All type-1 transmembrane glycoproteins are processed through the ER-Golgi pathway before signal peptide-driven cellular surface localization. The proposed polybasic site was functional (availability to proteases) when the S protein was on the virion for host cell entry. A fully posttranslationally processed S protein surface translocation could also provide a polybasic site to be processed by furin cleavage. However, there is no information on the availability or usability of the S protein's polybasic site by furin proteases in the cytoplasm before virus assembly. Thus, the NLS is functional in SARS-CoV-2-infected cells, and the polybasic site only functions during the viral entry step. The NLS is obviously functional in infected cells, and no furin cleavage at the polybasic site is necessary other than for viral entry. Our results confirmed that the S protein NLS motif was functional in SARS-CoV-2-infected cells. Although mutating the polybasic site (which also mutated the NLS) may impact viral S protein function in vitro, the result will not confirm or deny that one is more important than the other between the polybasic site and the NLS. While our result does provide direct evidence for the presence of the NLS motif and nuclear translocation of the S protein, our results do not confirm nor deny that the NSPR sequence has a natural origin. Instead, our results showed that the inserted

sequence NSPR was a functional NLS motif, which increased the intracellular distribution of the S protein, including novel nuclear translocation. The novel nuclear translocation of the SARS-CoV-2 S protein suggests that: 1. the nuclear translocation of the S protein reduces its surface expression, but whether it contributes to evading host immune recognition remains to be determined; and 2. the colocalization of the S protein with S mRNA suggests that the S protein has an RNA binding motif, which remains to be determined. One of the important ways of confirming a functional NLS motif is to use site-directed mutational analysis. Plasmid-driven transient expression of S protein in the human lung airway A549 cell line and primary normal human branchial epithelial cells showed robust S expression but was toxic to the cells. Therefore, the success of site-directed mutational analysis of the S protein in a transient expression system is doubtful and the characterization of NLS by a mutational analysis is yet to be determined. Thus, our novel findings emphasize further research on the NLS motif of the SARS-CoV-2 S protein.

One of the most important findings in our study was the simultaneous detection of the different spatial distributions of S protein and S mRNA at the single-molecule level in a single infected cell. We confirmed that S mRNA translocated into the nucleus by image analysis of the colocalization of S mRNA with nuclear staining. The SARS-CoV-2 N protein has already been shown to bind to RNA [46]. There was no information available confirming whether the S protein could bind to S mRNA for nuclear translocation. Our results revealed that S mRNA nuclear translocation was mediated by the S protein because S mRNA nuclear translocation was always associated with the S protein. For example, S mRNA colocalized with the S protein inside and outside the surface of the nucleus. Although the primer-probe was designed to target S mRNA, the SARS-CoV-2 positive-strand RNA genome (whole or partial) can be targeted by the same probe due to the sequence similarity between S mRNA and the whole or partial genome. Thus, our results

lack sufficient detail contributing to the discussion of the controversial scientific topic of whether there is any possibility of SARS-CoV-2 genome integration into the host DNA [47, 48]. Additionally, one of the significant differences in the S protein sequences of SARS-CoV and SARS-CoV-2 is the pat7 NLS motif. Whether S protein expression by the current vaccine platforms causes suboptimal expression of S protein on the cell surface due to the NLS remains to be determined [49].

In conclusion, the SARS-CoV-2 S protein has a functional pat7 NLS "PRRARSV", that results in one out of four S proteins translocating into the nucleus in infected cells. S Protein appears to shuttle S mRNA (possibly the genome) into the nucleus as well. Thus, the NLS of the S protein may contribute to the evasion of the host immune response and is a novel pathogenic feature of SARS-CoV-2.

Materials and Methods

Cells and viruses. Primary normal human bronchial epithelial (NHBE) cells from healthy adults and high-risk adults (deidentified) were obtained from Dr. Kristina Bailey at the University of Nebraska Medical Center (UNMC) (Omaha, NE) under an approved material transfer agreement (MTA) between the University of North Dakota (UND) and UNMC (Omaha, NE). The protocol for obtaining cells was reviewed by the UNMC IRB and was determined to not constitute human subject research (#318-09-NH). In this study, we used cells from five donors: nonsmoker healthy adults (donors #1 and #2) and adult with chronic obstructive pulmonary disease (COPD) (donor #3). The protocols for subculturing primary NHBE cells were published previously [2, 43, 50]. SARS-CoV-2 (USA/WA-CDC-WA1/2020 isolate, GenBank accession no. MN985325; kindly provided by CDC), SARS-CoV (Urbani strain, GenBank accession no. AY278741; kindly

334 provided by Rocky Mountain Laboratories (RML), NIAID, NIH), and MERS-CoV (GenBank
335 accession no. NC_019843.3; kindly provided by the Department of Viroscience, Erasmus Medical
336 Center, Rotterdam, The Netherlands) were used for *in vitro* infections described below.

337

338

339 **In silico analysis.** We have used open-source web portals for different *in silico* analyses. 1.
340 Constraint-based alignment tool for multiple protein sequences (COBALT)
341 (https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi) was used for multiple sequence
342 alignment. 2. PSORT II (https://psort.hgc.jp/form2.html) was used for NLS prediction. 3. The
343 RPIseq web portal (http://pridb.gdcb.iastate.edu/RPISeq/) was used for RNA–protein interactions.

344

345 **Highly differentiated pseudostratified bronchial airway epithelium.** The protocols for
346 differentiating primary NHBE cells to form a pseudostratified bronchial airway epithelium were
347 published previously [2, 43, 50]. Briefly, Transwells (6.5 mm) with 0.4-µm-pore polyester
348 membrane inserts (Corning Inc.) were coated with PureCol (Advanced BioMatrix) for 20 min
349 before cell seeding. NHBE cells (5x10^4) suspended in 100 µl of complete airway epithelial cell
350 (cAEC) medium [AEC medium (Promocell) + SupplementMix (Promocell) + 1%
351 penicillin–streptomycin (V/V) (Thermo Fisher Scientific) + 0.5% amphotericin B (V/V) (Thermo
352 Fisher Scientific)] were seeded in the upper chamber of the Transwell. Then, 500 µl of cAEC
353 medium was added to the lower chamber of the Transwell. When the cells formed a confluent
354 layer on the Transwell insert, the cAEC medium was removed from the upper chamber, and in the
355 lower chamber, the cAEC medium was replaced with complete ALI medium [PneumaCult-ALI

356 basal medium (Stemcell Technologies Inc.) + with the required supplements (Stemcell
357 Technologies) + 2% penicillin–streptomycin (V/V) + 1% amphotericin B (V/V)]. The complete
358 ALI medium in the lower chamber was changed every day. The upper chamber was washed with
359 1x Dulbecco's phosphate-buffered saline (DPBS) (Thermo Fisher Scientific) once per week
360 initially but more frequently when more mucus was observed during later days. All cells were
361 differentiated for at least four weeks at 37°C in a 5% CO_2 incubator. We observed motile cilia in
362 the differentiated airway epithelium similar to previously described [50].

363

364 **Viral infection.** All viral infection experiments were conducted in the high biocontainment facility
365 at RML, NIAID, NIH, Hamilton, MT. After approximately 3 weeks, the differentiated airway
366 epithelium on Transwells was shipped to RML in an optimized transportation medium [2, 50], and
367 the recovered cells were maintained in complete ALI medium for approximately one week before
368 infection. For infection, the airway epithelium on Transwells was washed with 200 µl of 1x PBS
369 to remove mucus and were infected on the apical site with SARS-CoV-2, MERS-CoV, or SARS-
370 CoV at a MOI of 0.1 in 100 µl 1x PBS for 1 hour (at 37°C with 5% CO_2). For mock infection, the
371 Transwells were similarly incubated with 100 µl 1x PBS without virus. The viral inoculum was
372 then removed, and the epithelium on the Transwell was washed twice with 200 µl of 1x PBS.
373 Complete ALI medium (1000 µl) was added to the lower chamber of each Transwell, and the upper
374 chamber was kept empty. Mock-infected and virus-infected Transwells were incubated for 4 days
375 at 37°C in an incubator with 5% CO_2 [2].

376

377 **Paraformaldehyde (PFA) fixation and paraffin embedding.** At 4 days postinfection (DPI), 200
378 µl of 1x PBS was added to the apical site of the Transwell for washing before PFA fixation. In the
379 basal side of the Transwell inserts was 200 µl of 1x PBS. For PFA fixation, 200 µl of 4% PFA
380 (Polysciences) was added to the upper chamber of the Transwells and incubated for 30 min, and
381 the Transwells were further maintained overnight in 4% PFA prior to removal from high
382 biocontainment. The PFA fixation protocol was approved as an inactivation method for
383 coronaviruses by the RML Institutional Biosafety Committee. The PFA-fixed airway epithelium
384 was paraffin-embedded and sectioned at a thickness of 5 µm for slide preparation as previously
385 described [43].

386

387 **Simultaneous detection of S mRNA and S protein.** Slides with 5 µm sections were first
388 deparaffinized by incubation in a Coplin jar as follows: 1. Histo-Clear for 5 min, two times; 2.
389 100% ethanol for 5 min, three times; 3. 95% ethanol for 5 min, 4. 70% ethanol for 5 min, and 5.
390 distilled water for 5 min. The deparaffinized slides were immediately incubated in 0.5% Triton X-
391 100 in 1x PBS for 30 min. The slides were washed three times with 1X PBST (1x PBS with Tween
392 20) or 1x PBS for 5 min. A hydrophobic barrier was drawn around the 5 µm section on the slides
393 by using an Immedge Hydrophobic Barrier Pen. To reduce nonspecific antibody binding, the
394 section was blocked with 10% goat serum (Vector Laboratories) in 1x PBST for 2 hours at 4°C.
395 The slides were then incubated with viral protein-specific primary antibody solution in 1x PBST
396 (e.g., SARS-CoV/SARS-CoV-2 S protein-specific rabbit polyclonal antibody at a 1:100 dilution,
397 SARS-CoV/SARS-CoV-2 N protein-specific mouse monoclonal antibody at a 1:100 dilution, or
398 MERS-CoV N protein-specific mouse monoclonal antibody at a 1:100 dilution) overnight at 4°C.
399 The slides were then incubated with the corresponding secondary antibody solution (anti-mouse

or anti-rabbit AF488 or AF647, Thermo Fisher Scientific) in 1x PBST for 2 hours at room temperature. We then stained the nuclei with DAPI reagent (Advanced Cell Diagnostics) or used RNAscope multiplex V2 to detect SARS-CoV-2 S mRNA (Probe V-nCoV2019-S) according to the manufacturer's instructions (Advanced Cell Diagnostics). The sections were mounted on Tech-Med microscope slides (Thomas Fisher Scientific) using ProLong-Gold antifade mounting medium (Thermo Fisher Scientific).

Imaging and image analysis. The images were taken under an Olympus FluoView laser scanning confocal microscope (Olympus FV3000) enabled with a 60X objective (Olympus), a Leica Stellaris confocal microscope (Leica Microsystem) using a 63x oil objective or a Leica DMI8 epifluorescence microscope (Leica Microsystem). The images were then deconvolved using Huygen Essential deconvolution software (Scientific Volume Imaging). The surface rendering function of Imaris image processing software (Oxford Instruments) was used. The images were also analyzed for spot-to-spot colocalization by Imaris. Where applicable, images taken under a Leica DMI8 microscope were processed using 3D deconvolution and 3D view modules in LASX software (Leica Microsystem). For figure preparation, Prism version 9 (GraphPad) and Adobe Photoshop (Creative Cloud) software were used.

Fig 1. Only the SARS-CoV-2 S protein had an NLS motif "PRRARSV" due to a novel sequence insertion. A. Full-length SARS-CoV-2 genome (nucleotide) (USA/WA-CDC-WA1/2020 isolate, GenBank accession no. MN985325) and open reading frames (ORF) are shown at the top. The SARS-CoV-2 S protein amino acid sequence was aligned with SARS-CoV (Urbani

422 strain, GenBank accession no. AY278741) by NCBI's constraint-based multiple alignment tool
423 COBALT [25], and the relative positions of four novel sequence insertions (ISs) are shown in the
424 S protein ORF as follows: IS1: "GTNGKTR", IS2: "YYHK", IS3: "HRSY", and IS4: "NSPR".
425 The fourth IS (NSPR) created a pat7 NLS "PRRARSV" in the S protein (shown in the red
426 rectangle). **B.** The S protein ORF sequences between SARS-CoV-2 and SARS-CoV were aligned,
427 and the Lined rectangles highlight the four novel insertions: IS1, IS2, IS3, and IS4. The IS4
428 "NSPR" created a pat7 NLS "PRRARSV" in the S protein (shown in the black rectangle).

429

430 **Fig 2. The intracellular distribution of S mRNA and S protein suggests nuclear translocation.**
431 Four-week **highly differentiated** pseudostratified airway epithelium was infected with SARS-
432 CoV-2 at a MOI of 0.1 for four days, paraformaldehyde-fixed, paraffin-embedded, and sectioned
433 at a thickness of 5 μm for immunohistochemistry (IHC) and slide preparation [43, 50]. A combined
434 protocol of RNAscope and IHC was used to simultaneously detect S mRNA and S protein in the
435 SARS-CoV-2-infected airway epithelium**.** S mRNA (red) was detected using a SARS-CoV-2 S
436 mRNA probe for RNAscope, and S protein (cyan) was detected by an S protein-specific rabbit
437 polyclonal antibody and a corresponding anti-rabbit secondary antibody for immunofluorescence
438 (IFA) analysis. The nucleus (blue) was detected by DAPI staining. The images were taken under
439 an Olympus confocal microscope using a 60x oil objective. The images represent multiple
440 independent technical replicates from two independent experiments with different donors
441 (experiment 1: donors 2 and 3 and experiment 2: donor 1). The scale bar is 10 μm.

442

443 **Fig 3. The nuclear translocation of S protein and S mRNA includes both the outer surface**
444 **and inside of the nucleus.** Separate slides (see Fig. 2) were imaged under a Leica Stellaris
445 confocal microscope (Leica) using a 63x oil objective. The images were then deconvolved using
446 Huygen Essential deconvolution software (Scientific Volume Imaging). Using the surface
447 rendering function of an image processing IMARIS software. **A.** S mRNA (red) on the nuclear
448 surface (top) and inside the nucleus (bottom). White arrows indicate S protein on the nuclear
449 surface (top image) or inside the nucleus (bottom image). **B.** S protein (green) on the nuclear
450 surface (top image) and inside the nucleus (bottom image). White arrows indicate S protein on the
451 nuclear surface (top image) or inside the nucleus (bottom image). **C.** The total distribution of S
452 mRNA and S protein in the cells. The data were obtained by combining multiple images from an
453 independent experiment. **D.** The total colocalization between S mRNA and S protein in the cells.
454 The data were obtained by combining multiple images from an independent experiment.

455

456 **Fig 4. Colocalization between S mRNA and S protein inside infected cells.** The images (see
457 Fig. 3) were analyzed by using the surface rendering and colocalization features of IMARIS. S
458 protein and S mRNA distribution and colocalization in the cytoplasm (top panel), on the nuclear
459 surface (middle panel) and inside the nucleus (bottom panel). The specific region of colocalization
460 is indicated by a white spot. Scale bar 0.5 μm.

461

462 **Fig 5. The nucleoproteins of SARS-CoV, MERS-CoV, and SARS-CoV-2 translocate into the**
463 **nucleus. A.** Four-week pseudostratified airway epithelium was infected with SARS-CoV-2 at a
464 MOI of 0.1 for four days, fixed, paraffin-embedded, and sectioned at a thickness of 5 μm for

immunohistochemistry slide preparation. Simultaneous detection of S mRNA (shown in red) and N protein (shown in cyan) on the same slide was performed by a combined detection protocol in RNAscope-based mRNA and immunofluorescence-based protein detection. An S mRNA-specific probe was used for RNAscope, and an N protein-specific rabbit polyclonal antibody and the corresponding anti-rabbit secondary antibody were used. The nucleus (shown in blue) was detected by DAPI staining. The images were taken under an Olympus confocal microscope using a 60x oil objective. The images represent multiple independent technical replicates from two independent, healthy donors (top row: donor #1 and bottom row: donor #2). The scale bar is 10 μm. **B.** Four-week pseudostratified airway epithelium was infected with SARS-CoV-2, SARS-CoV, or MERS-CoV at an MOI of 0.1 for four days. SARS-CoV-2 or SARS-CoV N protein was detected by an N protein-specific rabbit polyclonal antibody and corresponding anti-rabbit secondary antibody. Similarly, the MERS N protein was detected by MERS N protein-specific primary and corresponding secondary antibodies. The nucleus (shown in blue) was detected by DAPI staining. The images represent multiple independent technical replicates from one experiment (donor #1).

Fig S1. NLS prediction in the S protein of pathogenic coronaviruses. All categories of NLS motifs were searched in the S protein sequence using the web-based program PSORT II (https://psort.hgc.jp/form2.html) [27] for the S protein ORF amino acid sequence of SARS-CoV-2 (USA/WA-CDC-WA1/2020 isolate, GenBank accession no. MN985325) (Query 1), SARS-CoV (Urbani strain, GenBank accession no. AY278741) (Query 2), or MERS-CoV (GenBank accession no. NC_019843.3) (Query 3).

Fig S2. Detection of the nuclear translocation of S mRNA and **S protein.** Confocal images of SARS-CoV-2-infected airway epithelium (described in Fig 2) were analyzed for spot-to-spot colocalization using Imaris image analysis software (Oxford Instruments). The left panel shows the confocal images, the middle panel shows spot-to-spot colocalization, and the right panel shows merged confocal images and spot-to-spot colocalization. Spot-to-spot colocalization between the nucleus and S protein or S mRNA is indicated by a different color. The images represent multiple independent cross sections of the SARS-CoV-2-infected airway epithelium (from 3 independent donors).

Fig S3. The translocation of S mRNA and S protein includes both the inside and outer surface of the nucleus. From the images shown in Fig 3, the signals of S mRNA and S protein were plotted in the graph by Imaris image analysis software. The distance and intensity of all S mRNA or S protein from the nuclear surface (considered 0) were plotted. A negative value indicates that S mRNA or S protein resides inside the nucleus. The higher the negative value is, the farther the distance from the nuclear surface. In contrast, a positive value indicates that S mRNA or S protein resides on the nucleus surface and beyond in the cytoplasm. The higher the positive value is, the farther the distance from the nuclear surface.

Fig S4. The SARS-CoV S protein does not translocate into the nucleus. A four-week pseudostratified airway epithelium was infected with SARS-CoV at an MOI of 0.1 for four days, fixed, paraffin-embedded and sectioned at a thickness of 5 μm for immunohistochemistry slide preparation. S protein (shown in cyan) was detected by immunofluorescence-based protein

509 detection using SARS-CoV/SARS-CoV-2 S specific rabbit polyclonal primary antibody and anti-
510 rabbit secondary antibody. The confocal image was analyzed for spot-to-spot colocalization using
511 Imaris image analysis software. The left panel shows a confocal image, the middle panel shows
512 spot-to-spot colocalization, and the right panel shows merged confocal images and spot-to-spot
513 colocalization. The images represent multiple independent cross sections of SARS-CoV-infected
514 airway epithelium (at least two donors).

515

516 **Fig S5. S mRNA and S protein colocalization was spatially evident in all possible ways inside**
517 **the infected cell.** The confocal images shown in Figs. 3 & 4 were further visualized at a higher
518 magnification to detect S mRNA and S protein colocalization spatially. S protein and S mRNA
519 distribution and colocalization in the cytoplasm (top panel), on the nuclear surface (middle panel)
520 and inside the nucleus (bottom panel). The specific region of colocalization is indicated by a white
521 spot. The colors were made translucent to show colocalization. Scale bar 0.2 μm.

522

523 **Fig S6. NLS motif prediction in the N protein of pathogenic coronaviruses.** All categories of
524 NLS motifs were searched in the N protein sequence using the web-based program PSORT II
525 (https://psort.hgc.jp/form2.html) [27] for the N protein ORF of SARS-CoV-2 (USA/WA-CDC-
526 WA1/2020 isolate, GenBank accession no. MN985325) (Query 1), SARS-CoV (Urbani strain,
527 GenBank accession no. AY278741) (Query 2), or MERS-CoV (GenBank accession no.
528 NC_019843.3) (Query 3).

529

Fig S7. Nuclear translocation of the N protein of pathogenic coronaviruses. Four-week pseudostratified airway epithelium was infected with SARS-CoV-2, SARS-CoV, or MERS-CoV at an MOI of 0.1 for four days, fixed, paraffin-embedded and sectioned at a thickness of 5 μm for immunohistochemistry slide preparation. SARS-CoV-2 or SARS-CoV N protein (green) was detected by a SARS-CoV/SARS-CoV-2 N protein-specific antibody. Similarly, the MERS N protein (green) was detected by the MERS N protein-specific antibody. The nucleus (shown in blue) was detected by DAPI staining. The confocal images were analyzed for spot-to-spot colocalization. The left panel shows a confocal image, the middle panel shows spot-to-spot colocalization, and the right panel shows merged confocal images and spot-to-spot colocalization. Spot colocalization between the nucleus and N protein is indicated by a different color. The images represent multiple independent technical replicates from at least one independent experiment for one donor (donor #1).

Fig S8. NLS motif distribution in the N protein in different pathogenic coronaviruses. The sequences of the N protein of the SARS-CoV-2 N protein (nCoV-WA1-2020, GenBank accession no. MN985325), SARS-CoV N protein (Urbani Strain, GenBank accession no. AY278741), and MERS-CoV N protein (HCoV-EMC/2012, GenBank accession no. NC_019843) by NCBI's constraint-based multiple alignment tool COBALT [25]. All categories of NLS motifs are shown in the colored rectangle box: pat4: green; pat7: blue; bipartite 1: black; bipartite 2: orange.

S1 dataset 1. Prediction of SARS-CoV-2 S protein and genome interaction

S2 dataset 2. Prediction of SARS-CoV-2 N protein and genome interaction

552

553 **Acknowledgments.** We are thankful to the MARC U-STAR program at UND for supporting the
554 undergraduate students Sarah Sattar and Kailey Jerome. We are grateful to Dr. Jaspreet K. Osan
555 for helping with primary cell culture work. We are also grateful to Heinz Feldmann of the
556 Laboratory of Virology, NIAID, NIH for support with material and infections. In addition, we
557 thank the Microscopy Core (UND, Grand Forks), which is funded by NIH P20GM103442, of the
558 INBRE program for providing access to an Olympus FV300 confocal microscope. Histological
559 services were provided by the UND Histology Core, which is supported by the NIH/NIGMS
560 awards P20GM113123, U54G M128729, and UND SMHS funds. We also thank the Imaging Core
561 (UND, Grand Forks), which is funded by NIH P20GM113123, NIH U54GM128729, and
562 UNDSMHS funds, for IMARIS image analyses. This work was funded by the NIH/NIGMS
563 awards P20GM113123 and T34GM122835, VA grant 101-BX005413 and partially by the
564 Intramural Research Program, NIAID, NIH. The content is solely the responsibility of the authors
565 and does not necessarily represent the official views of the NIH.

566

567 **Author contributions.** M.M. conceived the project and designed all the experiments. K.B.
568 provided the primary cells. F.F. performed the viral infection work. K.J., S.S. and M.M. performed
569 all staining for detection. M.M. and S.S. generated the microscopic images. J.K. and M.M.
570 processed and quantified images. M.M. analyzed (in silico) the viral genome and protein
571 sequences. M.M. prepared the figures and wrote and edited the manuscript.

572

573 **Conflicts of interest.** The authors declare no conflicts of interest.

Reference

1. Liu J, Xie W, Wang Y, Xiong Y, Chen S, Han J, et al. A comparative overview of COVID-19, MERS and SARS: Review article. Int J Surg. 2020;81:1-8. Epub 2020/07/31. doi: 10.1016/j.ijsu.2020.07.032. PubMed PMID: 32730205; PubMed Central PMCID: PMCPMC7382925.

2. Osan J, Talukdar SN, Feldmann F, DeMontigny BA, Jerome K, Bailey KL, et al. Goblet Cell Hyperplasia Increases SARS-CoV-2 Infection in Chronic Obstructive Pulmonary Disease. Microbiol Spectr. 2022:e0045922. Epub 20220713. doi: 10.1128/spectrum.00459-22. PubMed PMID: 35862971.

3. Sola I, Almazan F, Zuniga S, Enjuanes L. Continuous and Discontinuous RNA Synthesis in Coronaviruses. Annu Rev Virol. 2015;2(1):265-88. doi: 10.1146/annurev-virology-100114-055218. PubMed PMID: 26958916; PubMed Central PMCID: PMCPMC6025776.

4. Pasternak AO, Spaan WJM, Snijder EJ. Nidovirus transcription: how to make sense...? J Gen Virol. 2006;87(Pt 6):1403-21. doi: 10.1099/vir.0.81611-0. PubMed PMID: 16690906.

5. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. Nat Rev Microbiol. 2009;7(6):439-50. doi: 10.1038/nrmicro2147. PubMed PMID: 19430490; PubMed Central PMCID: PMCPMC2830095.

6. Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S, et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. Nature. 2020;583(7816):469-72. Epub 20200514. doi: 10.1038/s41586-020-2332-7. PubMed PMID: 32408336.

7. Finkel Y, Mizrahi O, Nachshon A, Weingarten-Gabbay S, Morgenstern D, Yahalom-Ronen Y, et al. The coding capacity of SARS-CoV-2. Nature. 2021;589(7840):125-30. Epub 20200909. doi: 10.1038/s41586-020-2739-1. PubMed PMID: 32906143.

597 8. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, et al. SARS-
598 CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven
599 Protease Inhibitor. Cell. 2020;181(2):271-80 e8. Epub 20200305. doi: 10.1016/j.cell.2020.02.052.
600 PubMed PMID: 32142651; PubMed Central PMCID: PMCPMC7102627.
601 9. Jia HP, Look DC, Shi L, Hickey M, Pewe L, Netland J, et al. ACE2 receptor expression
602 and severe acute respiratory syndrome coronavirus infection depend on differentiation of human
603 airway epithelia. J Virol. 2005;79(23):14614-21. doi: 10.1128/JVI.79.23.14614-14621.2005.
604 PubMed PMID: 16282461; PubMed Central PMCID: PMCPMC1287568.
605 10. Raj VS, Mou H, Smits SL, Dekkers DH, Muller MA, Dijkman R, et al. Dipeptidyl
606 peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature.
607 2013;495(7440):251-4. doi: 10.1038/nature12005. PubMed PMID: 23486063; PubMed Central
608 PMCID: PMCPMC7095326.
609 11. Conceicao C, Thakur N, Human S, Kelly JT, Logan L, Bialy D, et al. The SARS-CoV-2
610 Spike protein has a broad tropism for mammalian ACE2 proteins. PLoS Biol.
611 2020;18(12):e3001016. Epub 2020/12/22. doi: 10.1371/journal.pbio.3001016. PubMed PMID:
612 33347434; PubMed Central PMCID: PMCPMC7751883.
613 12. Martinez-Flores D, Zepeda-Cervantes J, Cruz-Resendiz A, Aguirre-Sampieri S, Sampieri
614 A, Vaca L. SARS-CoV-2 Vaccines Based on the Spike Glycoprotein and Implications of New
615 Viral Variants. Front Immunol. 2021;12:701501. Epub 20210712. doi:
616 10.3389/fimmu.2021.701501. PubMed PMID: 34322129; PubMed Central PMCID:
617 PMCPMC8311925.
618 13. Harrow J, Nagy A, Reymond A, Alioto T, Patthy L, Antonarakis SE, et al. Identifying
619 protein-coding genes in genomic sequences. Genome Biol. 2009;10(1):201. Epub 20090130. doi:

620 10.1186/gb-2009-10-1-201. PubMed PMID: 19226436; PubMed Central PMCID:
621 PMCPMC2687780.
622 14. Gussow AB, Auslander N, Faure G, Wolf YI, Zhang F, Koonin EV. Genomic determinants
623 of pathogenicity in SARS-CoV-2 and other human coronaviruses. Proceedings of the National
624 Academy of Sciences. 2020;117(26):15193-9.
625 15. Hicks GR, Raikhel NV. Protein import into the nucleus: an integrated view. Annual review
626 of cell and developmental biology. 1995;11(1):155-88.
627 16. Robbins J, Dilwortht SM, Laskey RA, Dingwall C. Two interdependent basic domains in
628 nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting
629 sequence. Cell. 1991;64(3):615-23.
630 17. Rowland RR, Chauhan V, Fang Y, Pekosz A, Kerrigan M, Burton MD. Intracellular
631 localization of the severe acute respiratory syndrome coronavirus nucleocapsid protein: absence
632 of nucleolar accumulation during infection and after expression as a recombinant protein in vero
633 cells. J Virol. 2005;79(17):11507-12. Epub 2005/08/17. doi: 10.1128/JVI.79.17.11507-
634 11512.2005. PubMed PMID: 16103202; PubMed Central PMCID: PMCPMC1193611.
635 18. Zhang J, Cruz-Cosme R, Zhuang MW, Liu D, Liu Y, Teng S, et al. A systemic and
636 molecular study of subcellular localization of SARS-CoV-2 proteins. Signal Transduct Target
637 Ther. 2020;5(1):269. Epub 2020/11/19. doi: 10.1038/s41392-020-00372-8. PubMed PMID:
638 33203855; PubMed Central PMCID: PMCPMC7670843.
639 19. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of
640 SARS-CoV-2. Nat Med. 2020;26(4):450-2. Epub 2020/04/15. doi: 10.1038/s41591-020-0820-9.
641 PubMed PMID: 32284615; PubMed Central PMCID: PMCPMC7095063.

642 20. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function,
643 and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020;183(6):1735. Epub
644 2020/12/12. doi: 10.1016/j.cell.2020.11.032. PubMed PMID: 33306958; PubMed Central
645 PMCID: PMCPMC7833104.
646 21. Zhang C, Zheng W, Huang X, Bell EW, Zhou X, Zhang Y. Protein Structure and Sequence
647 Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique
648 Similarity between Its Spike Protein Insertions and HIV-1. J Proteome Res. 2020;19(4):1351-60.
649 Epub 2020/03/24. doi: 10.1021/acs.jproteome.0c00129. PubMed PMID: 32200634; PubMed
650 Central PMCID: PMCPMC7099673.
651 22. Xie Y, Karki CB, Du D, Li H, Wang J, Sobitan A, et al. Spike Proteins of SARS-CoV and
652 SARS-CoV-2 Utilize Different Mechanisms to Bind With Human ACE2. Front Mol Biosci.
653 2020;7:591873. Epub 2020/12/29. doi: 10.3389/fmolb.2020.591873. PubMed PMID: 33363207;
654 PubMed Central PMCID: PMCPMC7755986.
655 23. Peacock TP, Goldhill DH, Zhou J, Baillon L, Frise R, Swann OC, et al. The furin cleavage
656 site in the SARS-CoV-2 spike protein is required for transmission in ferrets. Nat Microbiol.
657 2021;6(7):899-909. Epub 20210427. doi: 10.1038/s41564-021-00908-w. PubMed PMID:
658 33907312.
659 24. Ord M, Faustova I, Loog M. The sequence at Spike S1/S2 site enables cleavage by furin
660 and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV. Sci Rep.
661 2020;10(1):16944. Epub 20201009. doi: 10.1038/s41598-020-74101-0. PubMed PMID:
662 33037310; PubMed Central PMCID: PMCPMC7547067.

663 25. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple
664 protein sequences. Bioinformatics. 2007;23(9):1073-9. Epub 2007/03/03. doi:
665 10.1093/bioinformatics/btm076. PubMed PMID: 17332019.
666 26. Hu T, Liu Y, Zhao M, Zhuang Q, Xu L, He Q. A comparison of COVID-19, SARS and
667 MERS. PeerJ. 2020;8:e9725. Epub 20200819. doi: 10.7717/peerj.9725. PubMed PMID:
668 32879801; PubMed Central PMCID: PMCPMC7443081.
669 27. Nakai K, Horton P. PSORT: a program for detecting sorting signals in proteins and
670 predicting their subcellular localization. Trends Biochem Sci. 1999;24(1):34-6. Epub 1999/03/24.
671 doi: 10.1016/s0968-0004(98)01336-x. PubMed PMID: 10087920.
672 28. Tian S, Huang Q, Fang Y, Wu J. FurinDB: A database of 20-residue furin cleavage site
673 motifs, substrates and their associated drugs. Int J Mol Sci. 2011;12(2):1060-5. Epub 20110208.
674 doi: 10.3390/ijms12021060. PubMed PMID: 21541042; PubMed Central PMCID:
675 PMCPMC3083689.
676 29. Krysan DJ, Rockwell NC, Fuller RS. Quantitative characterization of furin specificity.
677 Energetics of substrate discrimination using an internally consistent set of hexapeptidyl
678 methylcoumarinamides. J Biol Chem. 1999;274(33):23229-34. doi: 10.1074/jbc.274.33.23229.
679 PubMed PMID: 10438496.
680 30. Ozawa M, Fujii K, Muramoto Y, Yamada S, Yamayoshi S, Takada A, et al. Contributions
681 of two nuclear localization signals of influenza A virus nucleoprotein to viral replication. J Virol.
682 2007;81(1):30-41. Epub 2006/10/20. doi: 10.1128/JVI.01434-06. PubMed PMID: 17050598;
683 PubMed Central PMCID: PMCPMC1797272.
684 31. Boisvert M, Bouchard-Levesque V, Fernandes S, Tijssen P. Classic nuclear localization
685 signals and a novel nuclear localization motif are required for nuclear transport of porcine

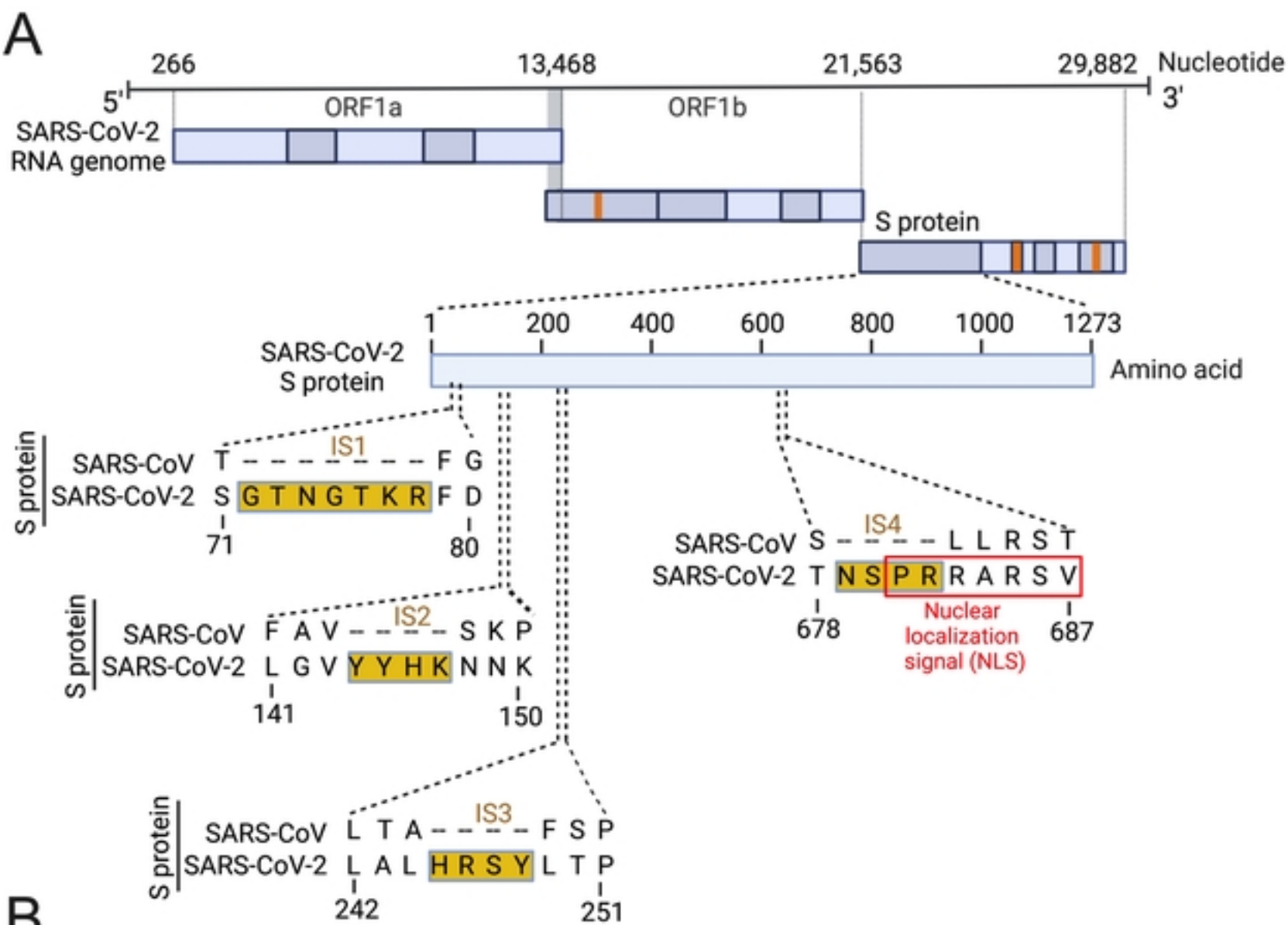
686 parvovirus capsid proteins. J Virol. 2014;88(20):11748-59. Epub 2014/08/01. doi:
687 10.1128/JVI.01717-14. PubMed PMID: 25078698; PubMed Central PMCID: PMCPMC4178750.
688 32. Vargas DY, Raj A, Marras SA, Kramer FR, Tyagi S. Mechanism of mRNA transport in
689 the nucleus. Proc Natl Acad Sci U S A. 2005;102(47):17008-13. Epub 2005/11/15. doi:
690 10.1073/pnas.0505580102. PubMed PMID: 16284251; PubMed Central PMCID:
691 PMCPMC1287982.
692 33. Whittaker GR, Helenius A. Nuclear import and export of viruses and virus genomes.
693 Virology. 1998;246(1):1-23. Epub 1998/07/10. doi: 10.1006/viro.1998.9165. PubMed PMID:
694 9656989.
695 34. Addetia A, Lieberman NAP, Phung Q, Hsiang TY, Xie H, Roychoudhury P, et al. SARS-
696 CoV-2 ORF6 Disrupts Bidirectional Nucleocytoplasmic Transport through Interactions with Rae1
697 and Nup98. mBio. 2021;12(2). Epub 2021/04/15. doi: 10.1128/mBio.00065-21. PubMed PMID:
698 33849972; PubMed Central PMCID: PMCPMC8092196.
699 35. Cubuk J, Alston JJ, Incicco JJ, Singh S, Stuchell-Brereton MD, Ward MD, et al. The
700 SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. Nat
701 Commun. 2021;12(1):1936. Epub 2021/03/31. doi: 10.1038/s41467-021-21953-3. PubMed
702 PMID: 33782395; PubMed Central PMCID: PMCPMC8007728.
703 36. Dinesh DC, Chalupska D, Silhan J, Koutna E, Nencka R, Veverka V, et al. Structural basis
704 of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. PLoS Pathog.
705 2020;16(12):e1009100. Epub 2020/12/03. doi: 10.1371/journal.ppat.1009100. PubMed PMID:
706 33264373; PubMed Central PMCID: PMCPMC7735635.
707 37. Timani KA, Liao Q, Ye L, Zeng Y, Liu J, Zheng Y, et al. Nuclear/nucleolar localization
708 properties of C-terminal nucleocapsid protein of SARS coronavirus. Virus Res. 2005;114(1-2):23-

709 34. Epub 2005/07/05. doi: 10.1016/j.virusres.2005.05.007. PubMed PMID: 15992957; PubMed
710 Central PMCID: PMCPMC7114095.
711 38. Yang Y, Zhang L, Geng H, Deng Y, Huang B, Guo Y, et al. The structural and accessory
712 proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus
713 (MERS-CoV) are potent interferon antagonists. Protein Cell. 2013;4(12):951-61. Epub
714 2013/12/10. doi: 10.1007/s13238-013-3096-8. PubMed PMID: 24318862; PubMed Central
715 PMCID: PMCPMC4875403.
716 39. Wu KE, Fazal FM, Parker KR, Zou J, Chang HY. RNA-GPS Predicts SARS-CoV-2 RNA
717 Residency to Host Mitochondria and Nucleolus. Cell Syst. 2020;11(1):102-8 e3. Epub 2020/07/17.
718 doi: 10.1016/j.cels.2020.06.008. PubMed PMID: 32673562; PubMed Central PMCID:
719 PMCPMC7305881.
720 40. Muppirala UK, Honavar VG, Dobbs D. Predicting RNA-protein interactions using only
721 sequence information. BMC Bioinformatics. 2011;12:489. Epub 2011/12/24. doi: 10.1186/1471-
722 2105-12-489. PubMed PMID: 22192482; PubMed Central PMCID: PMCPMC3322362.
723 41. Pradhan P, Pandey AK, Mishra A, Gupta P, Tripathi PK, Menon MB, et al. Uncanny
724 similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag. BioRxiv.
725 2020.
726 42. Gibson CA, Daniels RS, Oxford JS, McCauley JW. Sequence analysis of the equine H7
727 influenza virus haemagglutinin gene. Virus Res. 1992;22(2):93-106. Epub 1992/02/01. doi:
728 10.1016/0168-1702(92)90037-a. PubMed PMID: 1566601.
729 43. Osan JK, DeMontigny BA, Mehedi M. Immunohistochemistry for protein detection in
730 PFA-fixed paraffin-embedded SARS-CoV-2-infected COPD airway epithelium. STAR Protoc.

731 2021;2(3):100663. Epub 2021/07/13. doi: 10.1016/j.xpro.2021.100663. PubMed PMID:
732 34250510; PubMed Central PMCID: PMCPMC8259228.
733 44. Parent LJ. New insights into the nuclear localization of retroviral Gag proteins. Nucleus.
734 2011;2(2):92-7. Epub 2011/07/09. doi: 10.4161/nucl.2.2.15018. PubMed PMID: 21738831;
735 PubMed Central PMCID: PMCPMC3127090.
736 45. DaPoian AT, Gomes AM, Oliveira RJN, Silva JL. Migration of vesicular stomatitis virus
737 glycoprotein to the nucleus of infected cells. Proceedings of the National Academy of Sciences of
738 the United States of America. 1996;93(16):8268-73. doi: 10.1073/pnas.93.16.8268. PubMed
739 PMID: WOS:A1996VB32500022.
740 46. Schmidt N, Lareau CA, Keshishian H, Ganskih S, Schneider C, Hennig T, et al. The SARS-
741 CoV-2 RNA-protein interactome in infected human cells. Nat Microbiol. 2021;6(3):339-53. Epub
742 2020/12/23. doi: 10.1038/s41564-020-00846-z. PubMed PMID: 33349665; PubMed Central
743 PMCID: PMCPMC7906908.
744 47. Zhang L, Richards A, Barrasa MI, Hughes SH, Young RA, Jaenisch R. Reverse-transcribed
745 SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed
746 in patient-derived tissues. Proc Natl Acad Sci U S A. 2021;118(21). Epub 2021/05/08. doi:
747 10.1073/pnas.2105968118. PubMed PMID: 33958444; PubMed Central PMCID:
748 PMCPMC8166107.
749 48. Smits N, Rasmussen J, Bodea GO, Amarilla AA, Gerdes P, Sanchez-Luque FJ, et al.
750 Human genome integration of SARS-CoV-2 contradicted by long-read sequencing. bioRxiv. 2021.
751 49. Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, et al. An
752 mRNA Vaccine against SARS-CoV-2 - Preliminary Report. N Engl J Med. 2020;383(20):1920-

753 31. Epub 2020/07/15. doi: 10.1056/NEJMoa2022483. PubMed PMID: 32663912; PubMed
754 Central PMCID: PMCPMC7377258.
755 50. Osan JK, Talukdar SN, Feldmann F, Ann DeMontigny B, Jerome K, Bailey KL, et al.
756 Goblet Cell Hyperplasia Increases SARS-CoV-2 Infection in COPD. bioRxiv. 2020. Epub
757 2020/11/18. doi: 10.1101/2020.11.11.379099. PubMed PMID: 33200131; PubMed Central
758 PMCID: PMCPMC7668735.

759



B

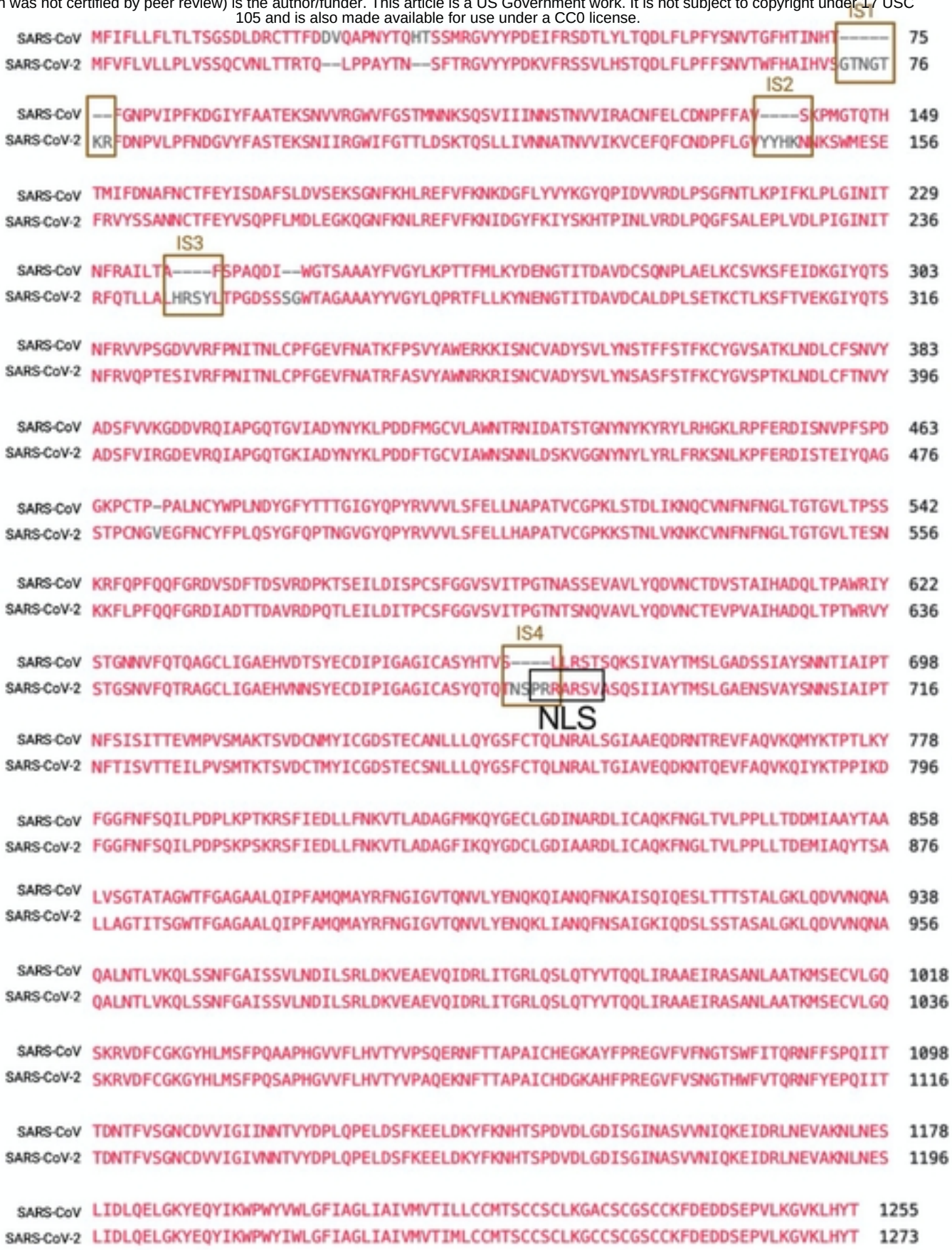


Fig 1

Fig 1

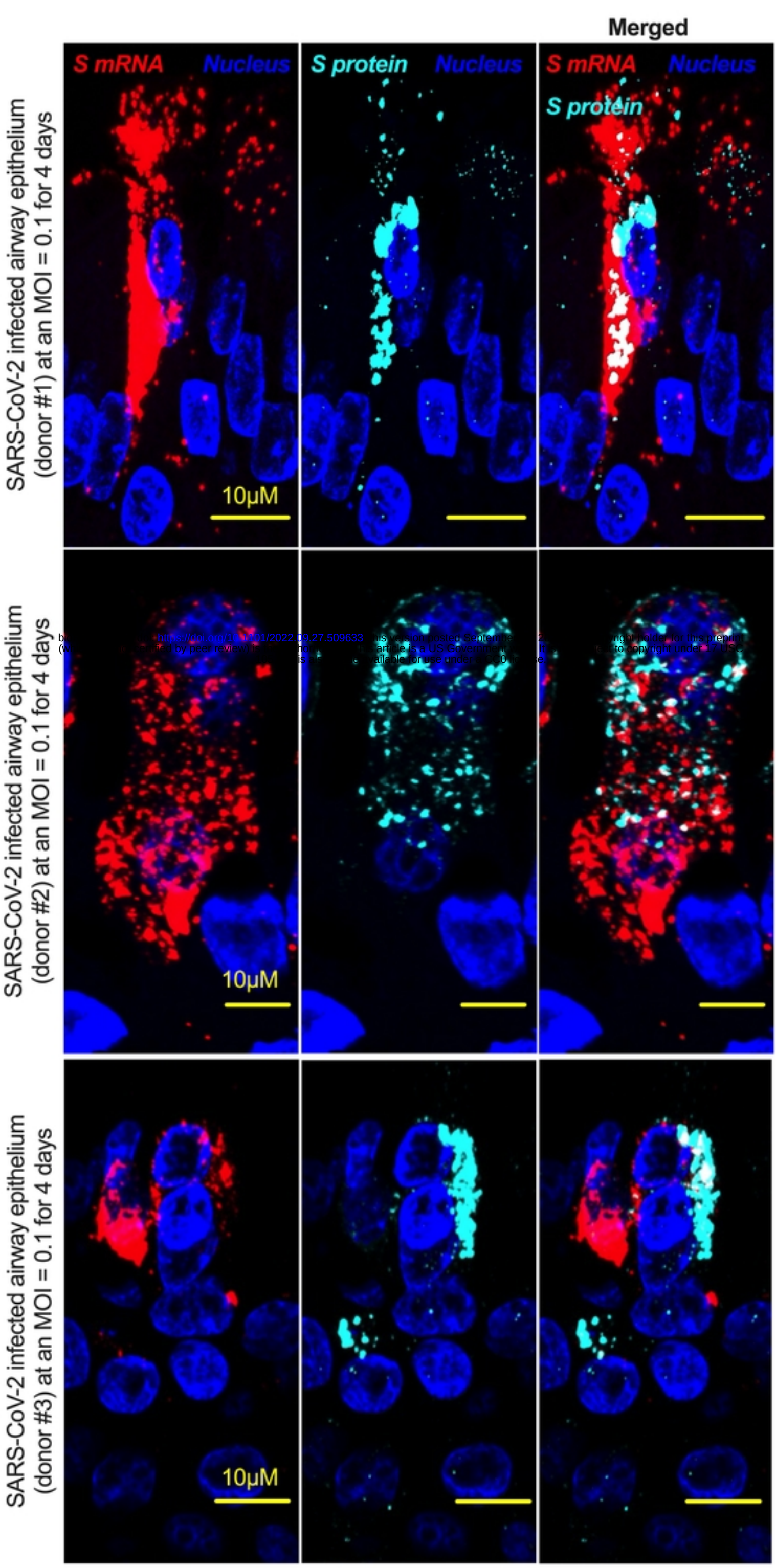


Fig 2

Fig 2

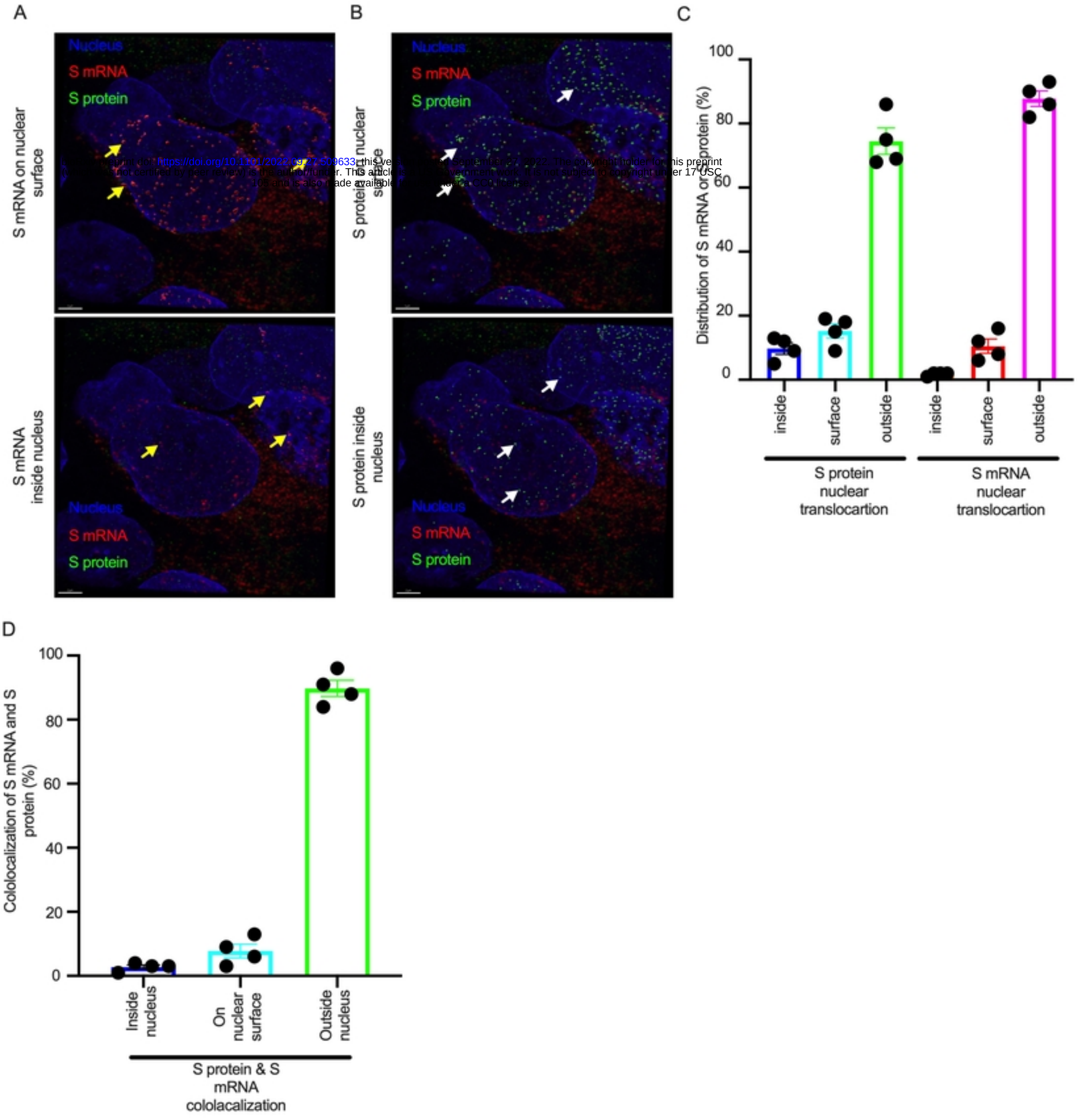


Fig 3

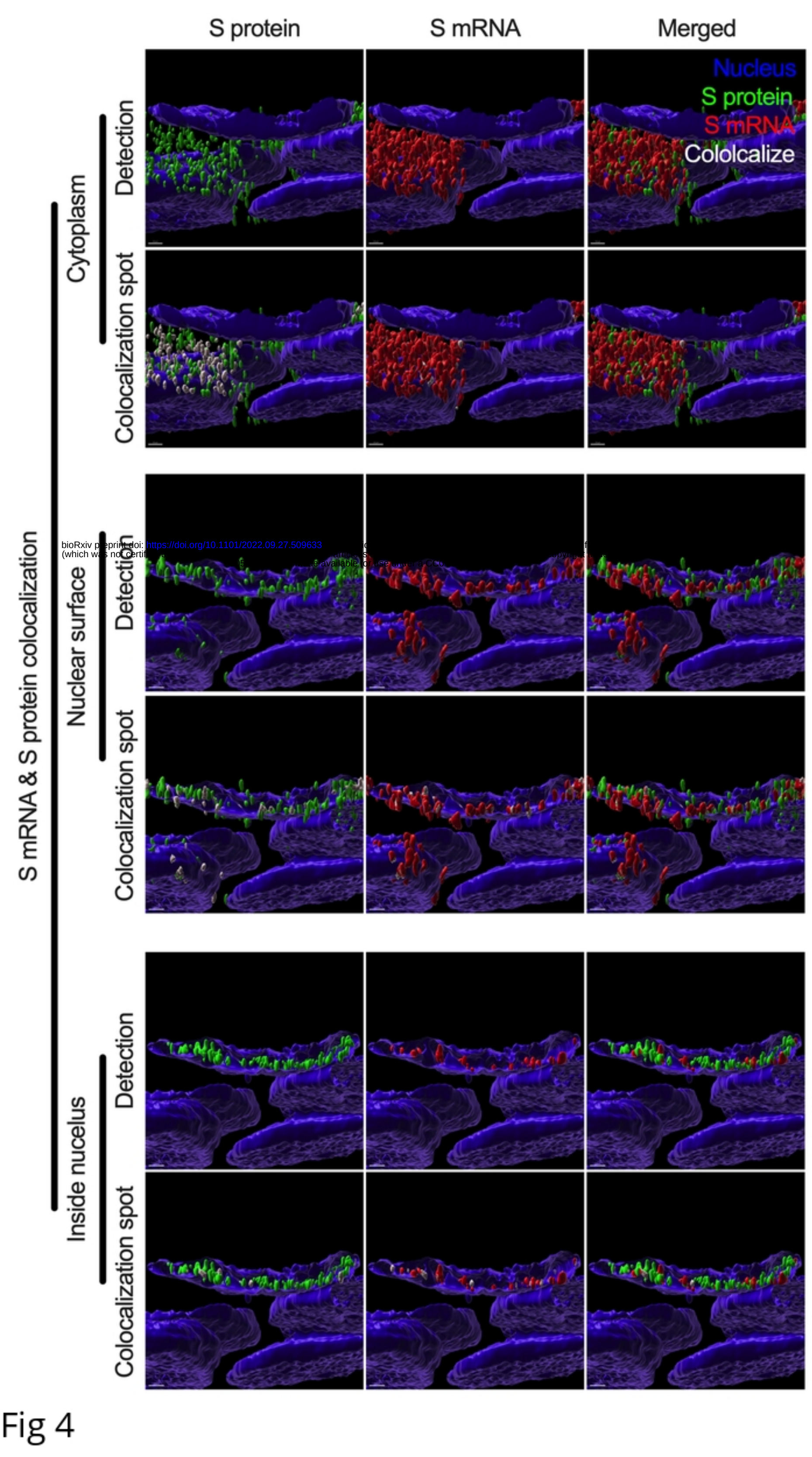
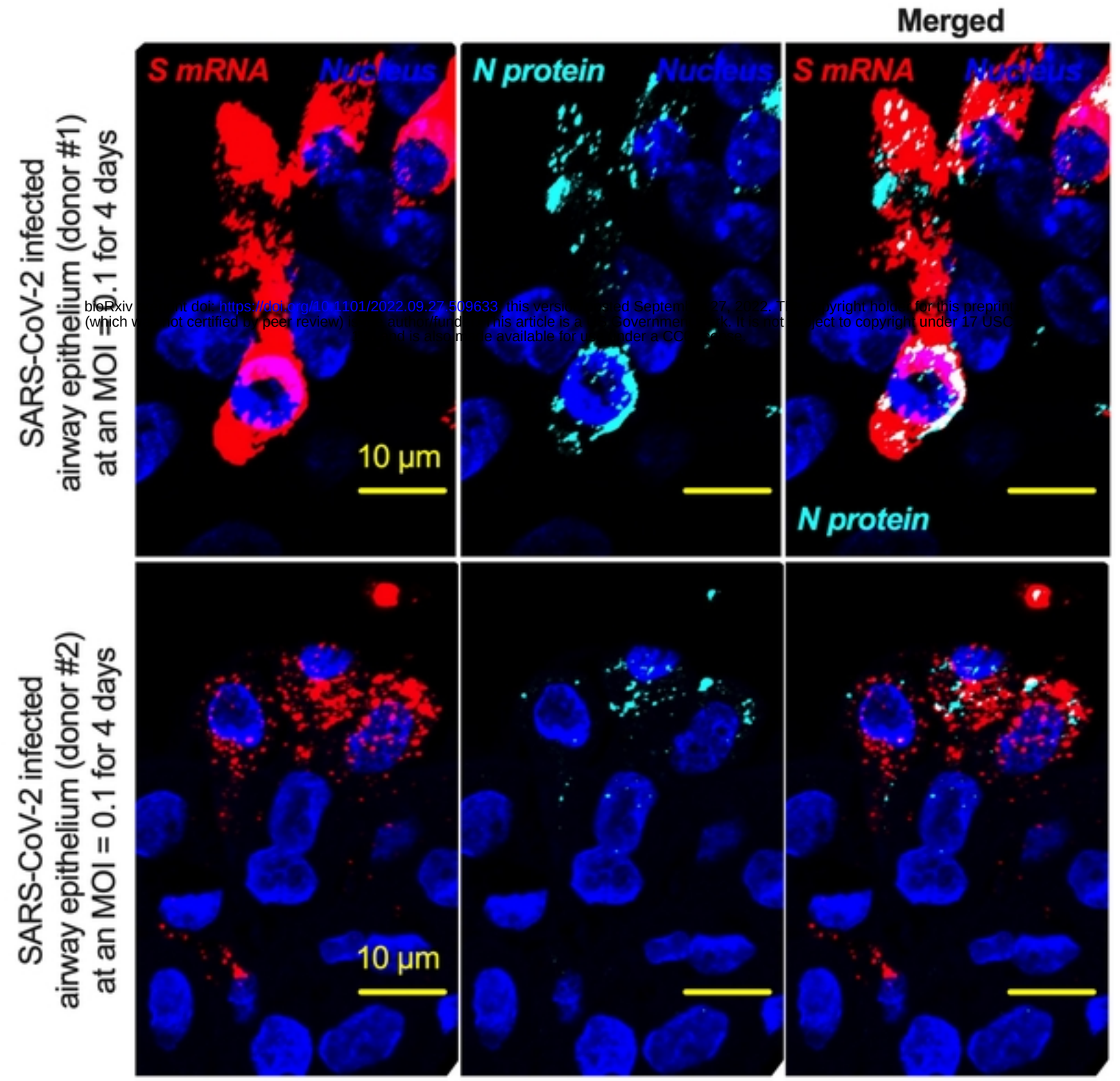


Fig 4

A



B

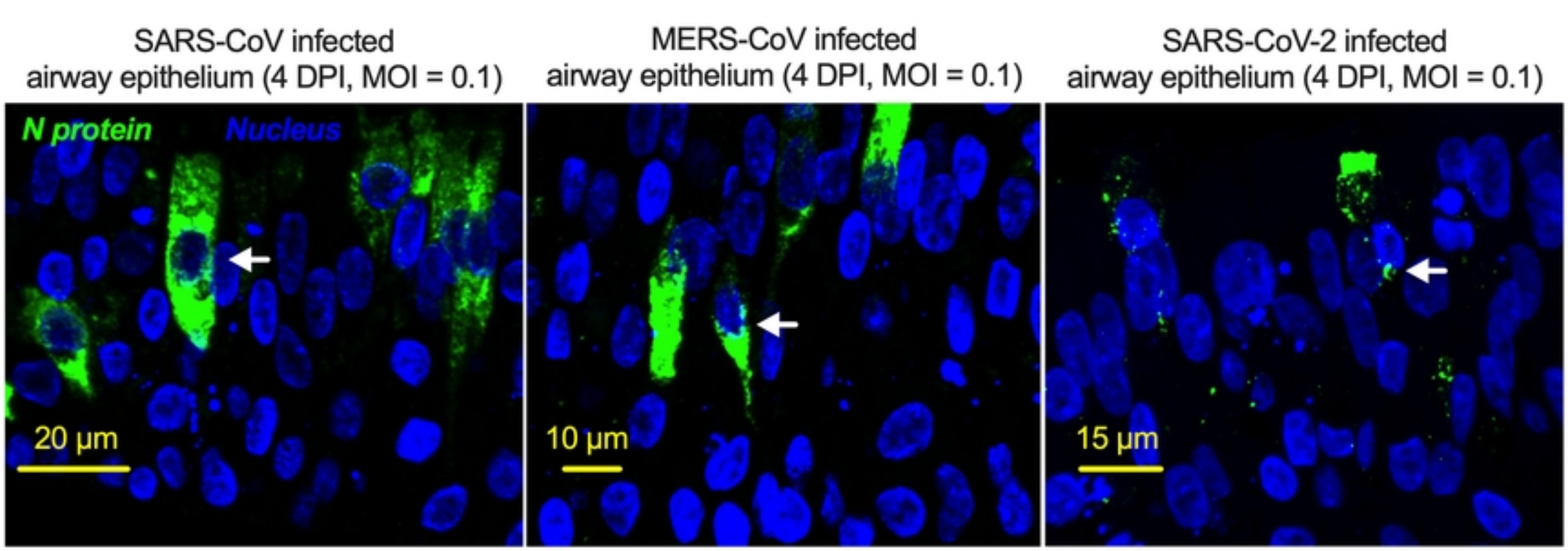


Fig 5