Sequences in the cytoplasmic tail of SARS-CoV-2 spike facilitate syncytia formation

Jerome Cattin-Ortolá[1,2], Lawrence Welch[1,2], Sarah L. Maslen[1], J. Mark Skehel[1], Guido Papa[1], Leo C. James[1] and Sean Munro[1]*

1: MRC Laboratory of Molecular Biology
Francis Crick Avenue
Cambridge CB2 0QH
UK

2: Denotes equal contribution
* Corresponding author: Email: sean@mrc-lmb.cam.ac.uk

Abstract

The spike (S) protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) binds the cell surface protein ACE2 to mediate fusion of the viral membrane with target cells[1-4]. S comprises a large external domain, a transmembrane domain (TMD) and a short cytoplasmic tail[5,6]. To elucidate the intracellular trafficking of S protein in host cells we applied proteomics to identify cellular factors that interact with its cytoplasmic tail. We confirm interactions with components of the COPI, COPII and SNX27/retromer vesicle coats, and with FERM domain actin regulators and the WIPI3 autophagy component. The interaction with COPII promotes efficient exit from the endoplasmic reticulum (ER), and although COPI-binding should retain S in the early Golgi system where viral budding occurs, the binding is weakened by a suboptimal histidine residue in the recognition motif. As a result, S leaks to the surface where it accumulates as it lacks an endocytosis motif of the type found in many other coronaviruses[7,8]. It is known that when at the surface S can direct cell:cell fusion leading to the formation of multinucleate syncytia[9-11]. Thus, the trafficking signals in the cytoplasmic tail of S protein indicate that syncytia formation is not an inadvertent by-product of infection but rather a key aspect of the replicative cycle of SARS-CoV-2 and potential cause of pathological symptoms.

Main text

Coronavirus virions are encapsulated by a lipid bilayer that contains a small set of membrane proteins including the S protein that binds to and fuses with target cells, and the M protein that recruits the viral genome into the virion[12-14]. Virions of SARS-CoV-2 and other betacoronviridae also contain the minor structural envelope (E) protein[15,16]. In infected cells, all three proteins are initially inserted into ER and are then trafficked to the ER-Golgi intermediate compartment (ERGIC) and the Golgi where they are glycosylated and, in the case of S, cleaved by Golgi enzymes[17]. Virions then form by budding into the membranes of the ERGIC and Golgi, and membrane-bound carriers then transfer the newly-formed virions to the surface for release from the cell[18,19]. Although SARS-CoV-2 and most other coronaviruses do not bud from the cell surface, some S protein is found on the surface of infected cells. Consistent with this, infected cells have been observed to fuse with neighbouring cells to form large multinucleate cells (syncytia)[9,10,20].

The traffic of membrane proteins through the secretory pathway is typically mediated by interactions between their cytoplasmic tails and the coat proteins that form the carriers that move proteins and lipids between compartments. To investigate the mechanisms by which S protein becomes distributed between the membranes of the host cell we sought cellular factors that recognised its cytoplasmic tail. The 37 residue tail of S was expressed as a GST fusion in bacteria and used for affinity chromatography of cytosol. Proteins that bound preferentially to the tail fusion rather than GST alone were identified by mass spectrometry (Fig. 1a, Extended Data Fig. 1a, and Supplementary Table 1). The highest hits included several vesicle coat proteins and three related proteins, ezrin, moesin and radixin, that link membrane proteins to actin. Ezrin was previously reported as hit in a yeast two-hybrid screen with the cytoplasmic tail of S from SARS, and suggested to have a role in restraining viral entry[21]. The vesicle coat proteins include subunits of the coatomer complex that forms COPI-coated vesicles, consistent with a previous report that coatomer binds to the closely related cytoplasmic tail of the S protein of SARS[22]. In addition to these known interactors, there was a strong enrichment of subunits of the COPII coat that forms ER-to-Golgi vesicles, and of the sorting nexin SNX27 and its associated retromer complex that together recycle proteins from endosomes to the cell surface[23]. The few other strong hits are not linked to membrane traffic and were not investigated further, with the exception of WIPI3/WDR45B which interacts with membranes to regulate autophagy[24].

To dissect the roles of the different coat proteins we mapped the regions that they bind on the 37 residue S protein tail (Fig. 1b,c). The tail comprises two distinct sections (Fig. 1d). The membrane proximal half (1234-1254) contains eight cysteines which are known to be palmitoylated in the equivalent region in SARS and other coronaviruses and, once modified, are likely to be embedded in the surface of the bilayer[25,26]. The distal half of the tail (1255-1273) lacks cysteines and so will project into the cytoplasm. Testing GST fusions to these two halves showed that all the interactors bound to the distal region with the exception of SNX27 which exclusively bound to the cysteine-rich region (Fig. 1c, Extended Data Figs. 1b and 2a). To map binding at higher resolution we tested tails with adjacent pairs of residues mutated to alanine (Fig. 1b). COPII binding was reduced by mutations in the acidic stretch DEDDSE that contains three copies of the di-acidic ER exit motif that binds to the Sec24 subunit of the coat[27,28]. In contrast, COPI binding required the residues in the C-terminal KXHXX motif that was also found to be required for this interaction in SARS S protein[22]. The FERM domain proteins required residues between the COPI and COPII binding sites, and SNX27 binding required residues in the N-terminal half of the tail nearest the TMD.

To further validate the interactions with SNX27, WIPI3 and the FERM domain proteins, we used recombinant proteins to test direct binding. Moesin is known to interact with plasma membrane proteins via its N-terminal FERM domain, and this part of the protein bound directly to the S protein tail, with residues 1261SEPV being essential (Fig. 2a). The autophagy regulator WIPI3, when expressed in E. coli, also bound directly to the membrane distal half of the tail (Fig. 2b). Recombinant SNX27 also bound to the tail in vitro, with residues 1238TSC next to the transmembrane domain being important, with the same residues found to be required for recruitment of SNX27 from cell lysate (Fig. 2c, Extended Data Fig. 2). SNX27 associates with retromer via the latter's VPS26 subunit, and we found that VPS26 is recruited to the tail of S protein by the addition of SNX27 indicating that the tail can bind SNX27 whilst it is in a complex with retromer (Fig. 2d). Interestingly, some tail mutants that lost SNX27 binding retained binding to retromer indicating that this complex can also bind elsewhere in the tail.

Mutation of residues required for SNX27 or moesin binding in full length S protein did not detectably alter its intracellular distribution or accumulation on the plasma membrane (Extended Data Fig. 3), indicating that these interactions do not have a role in cell surface delivery. Indeed, the binding of SNX27, although very efficient, is to the region of the tail that will be palmitoylated in host cells and so its in vivo significance remains unclear. We

thus examined the contribution of the COPI and COPII binding sites to the subcellular distribution of S protein. Mutation of the acidic residues in the COPII binding region greatly reduced cell surface expression, with S protein accumulating in the ER, indicating that these residues direct efficient egress of the newly-made S protein into the secretory pathway (Fig. 3a-c). The COPI binding region comprises KLHYT which differs somewhat from the canonical KXKXX or KKXX C-terminal COPI binding motif[29,30]. To test COPI binding we expressed a chimeric protein comprised of the S protein tail fused to the extracellular and TMD domains of the cell surface protein CD86. Immunoprecipitation of this chimera from transfected cells revealed co-precipitation with COPI and this was lost when lysine K1269 was mutated to alanine (Fig. 3d,e). When H1271 was altered to a canonical lysine, binding was substantially increased by both co-precipitation and affinity chromatography of cytosol, indicating that S protein has a suboptimal COPI binding site (Fig. 3e,f). In the above alanine scanning of the tail, mutation of the terminal residue T1273 to alanine was found to increase COPI binding (Fig. 1b), and this effect was recapitulated with the equivalent mutation in the CD86 chimera, suggesting the C-terminal threonine is a further feature of the tail that reduces its affinity for COPI (Fig. 3e). Extension of the tail with a C-terminal epitope tag also resulted in a loss of COPI binding indicating that, despite containing a histidine, the C-terminus is being recognised like a KXKXX motif that has to be at the C-terminus[30]. This also explains why the COPI interaction was not detected in recent proteomic searches for interaction partners of SARS-CoV-2 proteins as these used C-terminally tagged proteins[31,32].

Incorporation of the K1269A COPI binding site mutation into full length S protein caused, at most, only a small increase in the cell surface expression (Fig. 3g,h). In contrast, the H1271K and T1273A mutations that increased COPI binding both caused S protein to instead accumulate intracellularly with substantial co-localisation with the ER (Fig. 3g,h). Thus, the COPI binding site in S has conserved features that reduce its in vivo efficacy, and so allow it to reach the cell surface. Some other coronaviruses, including those that have histidine at the -3 position in the tail of S, have been found to be efficiently endocytosed if they reach the surface[7,8]. In these cases, endocytosis requires a tyrosine-containing motif that resembles the classic Yxxφ signal, but the S protein of SARS-CoV-2 lacks such a motif and consistent with this we found that it not only accumulated on the surface but did not show efficient endocytic uptake (Fig. 4a and Extended Data Fig. 3).

What might be the reason for S protein to accumulate at the cell surface? SARS-CoV-2, like other coronaviruses, buds into intracellular membranes and so S protein that has

reached the plasma membrane will not contribute to virion formation but it is in a position to cause infected cells to fuse to adjacent cells and so facilitate spread without virion release. We thus tested the effect of the mutations in the COPI binding site on the degree of cell fusion induced by S protein. 293T cells were transfected with a plasmid expressing S protein and mixed with Vero cells which endogenously express the S protein receptor ACE2, and cell fusion then followed with a fluorescent assay[33]. The K1269A mutant that prevents COPI binding caused a small but reproducible increase in cell fusion (Fig. 4b,c). In contrast, the H1271K mutation that binds better to COPI resulted in greatly reduced cell fusion, with the mutant S showing reduced levels of S1/S2 cleavage, consistent with it not moving beyond the early Golgi (Fig. 4d).

The data presented here argue that the S protein of SARS-CoV-2 has three features which facilitate accumulation on the plasma membrane. Firstly, a region containing di-acidic COPII binding motifs directs efficient exit from the ER. Secondly, the COPI-binding site is suboptimal which allows S protein to escape the Golgi apparatus. Consistent with this, the S protein of the coronavirus porcine epidemic diarrhea virus (PEDV) has a related C-terminal sequence (-KVHVQ) and was found to bind COPI with a much lower affinity than canonical KXKXX motifs (Fig. 4a)[34]. Finally, the S-protein of SARS-CoV-2 is not efficiently endocytosed, consistent with it lacking a tyrosine-containing motif of the sort that is found in many coronaviruses, including PEDV, and has been shown to either induce endocytosis or prevent movement beyond the Golgi depending on the virus (Fig. 4a)[7,8,35,36]. Our data do not provide insight into possible roles for the other interactions we have observed, and progress may require testing in different cells or in conditions of virus replication. Nonetheless, it is worth noting that FERM domain proteins anchor cell surface proteins to the actin cytoskeleton and thus direct their accumulation at specific regions[37].

Once at the cell surface, S protein is able to induce cell fusion and hence the formation of multinucleate syncytia. For other coronaviruses, including SARS, it is known that in virus-infected cells the M protein holds some S protein in the Golgi so as to direct its packaging in virions[13,22,38]. However, this does not preclude some S reaching the surface. Indeed, syncytia have been observed in cultured airway epithelial cells infected with SARS-CoV-2 in vitro, and in post mortem samples of patients who have died of COVID-19[9,10,39]. Thus, it is clear that some SARS-CoV-2 S can reach the surface during an infection. Our analysis of the tail of S protein suggests that this is not simply an irrelevant side reaction of virion production but rather a feature of S that is a consequence of the sequence of its cytoplasmic tail. Like other viruses, syncytium formation may be advantageous to SARS-

CoV-2 by allowing rapid and efficient spread between cells that is both rapid and can also evade immune surveillance[40]. Indeed, it has been argued that syncytia formation may be a common infection strategy amongst respiratory viruses[41]. It also seems quite conceivable that the formation of large syncytia increases viral pathogenicity by destablising airway epithelia or creating holes that are more challenging to repair than the loss of single infected cell.

Our findings thus argue that syncytia formation by SARS-CoV-2 should be viewed as a potential target for therapeutic strategies, especially as the process appears entirely dependent on the cell surface protease TMPRSS2 to make the activating S2' cleavage, whereas viral entry can be facilitated by either TMPRSS2 or lysosomal cathepsins[1,9].

Data Availability

Mass spectrometry data used in this study are summarised in Supplementary Table 1. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[42] partner repository with the dataset identifier PXDXXXXX (https://www.ebi.ac.uk/pride/archive/projects/PXDXXXXX). We made use of information in the UniProt data base https://www.uniprot.org/. All reagents generated by this study are available from the corresponding author on request.

Acknowledgments

We thank Manu Hegde for comments on the manuscript, John James and Natalya Leneva for reagents, Maria Daly for help with flow cytometry, and David Owen for advice on porcine diarrhea. Funding was from the Medical Research Council (MRC file reference number MC_U105178783).

Author Contributions

S.M. devised, and J.C.O., L.W. and S.M. planned, the study. J.C.O. and L.W carried out all the biochemistry, immunofluorescence and flow cytometry. S.L.M. performed the mass spectrometry, and S.L.M and J.M.S. processed the raw data. G.P and L.C.J. devised and performed the cell fusion assay. S.M. wrote the manuscript.

Competing Interests

The authors declare no competing interests.

Methods

Plasmids. Details of the plasmids used in this report can be accessed from Supplementary Table 2. Constructs used for bacterial expression of the cytoplasmic domains of the spike protein (S) of SARS-CoV-2: the sequence encoding the N-terminally GST-tagged cytoplasmic domain of S were cloned into the vector pGEX6p2 (GE Healthcare Life Sciences). For GST-S (1255-1273, pJC149) and GST-S (1237-1273, pJC150), the inserts and vector were amplified by PCR and assembled using Gibson cloning with the linker SDLEVLFQGPLGSPGIQ. For GST-S (1237-1254, pJC247) in pGEX6p2, sequence specific primers were annealed and inserted at the EcoRI and BamHI sites with the linker SDLEVLFQGPLGSPGIQ. For the alanine mutants of GST-S (pJC173 to pJC190, and pJC254 to pJC257) two fragments of pJC150 were amplified by PCR using mutation specific primers and assembled by Gibson cloning. gBlocks and PCR amplified vector were assembled by Gibson cloning. GST-S construct expressing the cytoplasmic domain of SARS-CoV (1219-1255) was generated by mutating GST-S from SARS-CoV-2: two fragments of pJC150 were amplified by PCR using mutation specific primers and assembled by Gibson cloning.

Constructs used for bacterial expression of the described S protein interactors: full length human WIPI3 (NCBI: NM_019613.4, UniProt Q5MNZ6) in pOPT vector (from O. Perisic, MRC-LMB) with a C-terminal PGAGA linker and a 6xHis tag was generated through Gibson assembly of the PCR amplified insert (pJC248 was the template) and vector. Human moesin (1-300)-6xHis (UniProt P26038), and human SNX27-6xHis (528 amino acid version UniProt Q96L92-3) were cloned into pOPT vector with a C-terminal PGAGA linker and a 6xHis tag. Inserts were synthesised as gene fragments (IDT for moesin [1-300] and Genewiz for SNX27) with codons optimized for E. coli expression. gBlocks and PCR amplified vectors were assembled by Gibson cloning.

WIPI3-HA construct used for expression in mammalian cells (pJC248): a gBlock (IDT) encoding WIPI3-HA was assembled by Gibson cloning with PCR amplified pcDNA3.1+ (Clontech). ss-HA-CPD in pEGFP_N1 (pJC338): signal sequence (1-31) with the HA tag in the reverse primer, carboxypeptidase D (CPD) (32-1380) and vector were PCR amplified and the three fragments were assembled by Gibson cloning. CPD cDNA was from GenScript clone ID OHu10876, accession number NM_001304.4.

Full length S protein constructs used for expression in mammalian cells: a sequence encoding SARS-CoV-2 S was codon optimised for expression in mammalian cells and cloned into pcDNA3.1+ (modified to be compatible with the PiggyBac transposase system) using the restriction sites NheI and NotI. Where indicated, an HA

tag was inserted after the signal peptide by introduction into the forward primer, amplification by PCR and insertion into pcDNA3.1+. Key residues in the cytoplasmic tail of S were mutated as indicated by introducing mutations into primers, amplification of a small region at the 3' end of the gene and insertion using the restriction sites BstEII and NotI. For GFP-CD86 chimeric fusions, GFP and a GAGAGS linker was inserted immediately downstream of the signal peptide of CD86 (the cDNA from John James) using Gibson assembly and inserted into modified pcDNA3.1+ using the restriction sites NheI and NotI. DNA fragments containing a short luminal region of CD86 (with a membrane-proximal insertion of two aspartic acid residues) and the TMD of CD86 fused to different mutant forms of the cytoplasmic tail of S were either synthesised (Genewiz) or amplified by PCR. The short fragments were cloned into the GFP-CD86 vector using the restriction sites BbvCI and NotI. GFP, with a short C-terminal linker, was cloned into pcDNA3.1+ using the restriction sites NheI and NotI.

Mammalian cell culture. Human embryonic kidney 293T (ATCC, CRL-3216), U2OS (ATCC, HTB-96) and Vero (ATCC, CCL-81) cells were cultured in Dulbecco's modified Eagle's medium Glutamax (DMEM; Gibco) supplemented with 10% fetal calf serum (FCS) and penicillin/streptomycin at 37°C and 5% CO_2. Unless indicated otherwise, cells were transfected using polyethylenimine (PEI; Polyscience, 24765) dissolved in PBS to 1 mg/mL. The ratio of PEI (µL) to DNA (µg) used was 3:1; PEI was dissolved in Opti-Mem, incubated at room temperature for 5 minutes, DNA was added and incubated for another 15 to 20 minutes at room temperature before dropwise addition onto cells which had been seeded the day before. Cells were free of mycoplasma as determined by regular testing (MycoAlert, Lonza).

Protein expression in bacteria. GST (pGEX6p2 vector), the different versions of the GST-S, E and M constructs and the 6xHis-tagged interactors were expressed as follows: plasmids were transformed into E. coli BL21-CodonPlus (DE3)-RIL (Agilent, 230245). From an overnight starter culture, cells were grown in 2xTY medium containing 100 µg/mL ampicillin (or 50 µg/mL Kanamycin for 6xHis-VPS26) and 34 µg/mL chloramphenicol at 37ºC in a shaking incubator. When the culture reached OD_{600} = 0.6 - 0.8, the temperature was lowered to 16ºC, protein expression was induced with 100 µM of Isopropyl β-D-1-thiogalactopyranoside (IPTG), and incubated overnight. Bacteria cells were harvested by centrifugation at 4,000 x *g* at 4°C for 15 minutes and were mechanically resuspended on ice in lysis buffer containing 50 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 5 mM 2-mercaptoethanol, 1% Triton X-100, and supplemented with

protease inhibitor cocktail (cOmplete, Roche). Cells were lysed by sonication and the lysates were clarified by centrifugation at 20,000 x *g* at 4ºC for 15 minutes. Clarified lysates were flash frozen in liquid nitrogen and thawed as needed for the binding assays.

GST-pulldowns using 293T cell lysates. Pull downs for mass spectrometry**:** clarified lysates from 450 mL 2xTY cultures containing bacteria expressing recombinant GST, GST-S tails (product of pJC149, pJC150, pJC247) were thawed. 100 µL of glutathione Sepharose 4B bead slurry (GE17-0756-01) was washed twice with lysis buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 1mM EDTA, 5 mM 2-mercaptoethanol, 1% Triton X-100) by centrifugation at 100 *x g* for 1 minute at 4ºC and aspiration of the washing buffer. Clarified bacterial lysates were added to the Glutathione Sepharose beads and incubated at 4ºC for 1 hour on a tube roller. 293T cells (from four confluent T175 flasks per GST-tagged bait) were collected by scraping and lysed with lysis buffer supplemented with protease inhibitor cocktail (EDTA-free, cOmplete, Roche). The lysate was clarified by centrifugation for 5 minutes at 17,000 *x g* and pre-cleared with 100 µL of Glutathione Sepharose bead slurry per bait. Beads loaded with recombinant GST-tagged baits were washed once with ice-cold lysis buffer, once with lysis buffer supplemented with 500 mM NaCl, and once again with lysis buffer. Around 5% of the beads were kept aside as an input control and the remaining beads were incubated with the pre-cleared 293T cell lysate for 2-4 hours on a tube roller at 4ºC. Beads were washed twice with lysis buffer, transferred to 0.8 mL centrifuge columns (Pierce 89869B) and washed twice more. Columns were brought to room temperature and eluted 5 times with 100 µL of elution buffer (1.5 M NaCl in lysis buffer) by centrifugation at 100 x *g* for 1 minute; for the final elution the sample was centrifuged at 17,000 x *g* for 1 minute. Eluates were pooled together and concentrated down to around 75 µL using an Amicon Ultra 0.5 mL 3,000 NMWL centrifugal filter (Millipore UFC500324), supplemented with 25 µL of NuPage 4x LDS sample buffer (Invitrogen, NP0007) containing 100 mM DTT. 40% of the eluate was loaded on SDS PAGE gels (Invitrogen, XP04202) and stained with InstantBlue Coomassie Protein Stain (Abcam, ab119211). Each lane was cut into 8 gel slices for mass spectrometry analysis.

Pulldowns for western blotting: pulldowns were conducted as described for mass spectrometry with the exception that half of the amounts of clarified bacterial lysate, Glutathione Sepharose beads, and clarified 293T cell lysate were used. The amounts of eluates indicated on the figures were loaded on to SDS PAGE gels (Invitrogen, XP04205). GST-pulldown with 293T cells expressing WIPI3-HA: cells were transiently transfected with WIPI3-HA (pJC248). 48 h later, cells were collected by scraping, lysed

and used for the pulldown as described above with the exception that for each bait, cells from half of a T175 flask were used.

Mass spectrometry. Polyacrylamide gel slices (1-2 mm) were placed in a well of a 96-well microtitre plate and destained with 50% v/v acetonitrile and 50 mM ammonium bicarbonate, reduced with 10 mM DTT, and alkylated with 55 mM iodoacetamide. After alkylation, proteins were digested with 6 ng/µL trypsin (Promega, UK) overnight at 37 °C. The resulting peptides were extracted in 2% v/v formic acid, 2% v/v acetonitrile. The digests were analysed by nano-scale capillary LC-MS/MS using an Ultimate U3000 HPLC (ThermoScientific Dionex, San Jose, USA) to deliver a flow of approximately 300 nL/min. A C18 Acclaim PepMap100 5 µm, 100 µm x 20 mm nanoViper (ThermoScientific Dionex, San Jose, USA), trapped the peptides prior to separation on a C18 BEH130 1.7 µm, 75 µm x 250 mm analytical UPLC column (Waters, UK). Peptides were eluted with a 60-minute gradient of acetonitrile (2% to 80%). The analytical column outlet was directly interfaced via a nano-flow electrospray ionisation source, with a quadrupole Orbitrap mass spectrometer (Q-Exactive HFX, ThermoScientific, USA). MS data were acquired in data-dependent mode using a top 10 method, where ions with a precursor charge state of 1+ were excluded. High-resolution full scans (R=60,000, m/z 300-1800) were recorded in the Orbitrap followed by higher energy collision dissociation (HCD) (26 % Normalized Collision Energy) of the 10 most intense MS peaks. The fragment ion spectra were acquired at a resolution of 15,000 and dynamic exclusion window of 20s was applied. Raw data files from LC-MS/MS data were were processed using Proteome Discoverer v2.1 (Thermo Scientific), and then searched against a human protein database (UniProt KB, reviewed) using the Mascot search engine programme (Matrix Science, UK).

Database search parameters were set with a precursor tolerance of 10 ppm and a fragment ion mass tolerance of 0.2 Da. One missed enzyme cleavage was allowed and variable modifications for oxidized methionine, carbamidomethyl cysteine, pyroglutamic acid, phosphorylated serine, threonine and tyrosine were included. MS/MS data were validated using the Scaffold programme (Proteome Software Inc., USA). All data were additionally interrogated manually. The data presented was exported from Scaffold as total spectral counts with the protein threshold was set at 80%, the minimum number of peptides was set as 2 and the peptide threshold was set at 50%. the Lower Scoring Matches, and those of <5% probability were not shown.

Analysis of mass spectral intensities. All raw files were processed with MaxQuant v1.5.5.1 using standard settings and searched against the UniProt Human Reviewed KB

with the Andromeda search engine integrated into the MaxQuant software suite[43,44]. Enzyme search specificity was Trypsin/P for both endoproteinases. Up to two missed cleavages for each peptide were allowed. Carbamidomethylation of cysteines was set as fixed modification with oxidized methionine and protein N-acetylation considered as variable modifications. The search was performed with an initial mass tolerance of 6 ppm for the precursor ion and 0.5 Da for MS/MS spectra. The false discovery rate was fixed at 1% at the peptide and protein level. Statistical analysis was carried out using the Perseus module of MaxQuant[45]. Prior to statistical analysis, peptides mapped to known contaminants, reverse hits and protein groups only identified by site were removed. Only protein groups identified with at least two peptides, one of which was unique and two quantitation events were considered for data analysis. Each protein had to be detected in at least two out of the three replicates. Missing values were imputed by values simulating noise using the Perseus default settings. To calculate the p-values, two sample t-tests was performed.

Immunoblotting. Protein samples in 1x NuPage LDS sample buffer, 25 mM DTT (or 50 mM TCEP pH 7.0) were loaded on to SDS PAGE gels (Invitrogen, XP04205) and transferred to nitrocellulose membranes. Membranes were blocked in 5% (w/v) milk in PBS-T (PBS with 0.1% [v/v] Tween-20) for 1 hour, incubated overnight at 4°C with primary antibody in the same blocking solution, washed three times with PBS-T for 5 minutes, incubated with HRP- or Alexa Fluor (AF)-conjugated secondary antibody (where indicated) in 0.1% (w/v) milk in PBS-T for 1 hour and, washed three times with PBS-T for 5 minutes. Where indicated a primary anti-6xHis or anti-SNX27 antibody crosslinked to HRP was used. Blots stained with HRP-conjugated secondary antibodies were imaged using BioRad ChemiDoc imagers with chemiluminescence substrates: either SuperSignal West Pico PLUS (Thermo Scientific, 34577) or SuperSignal West Femto (Thermo Scientific, 34095) depending on the strength of the signal. Blots stained with AF-conjugated secondary antibodies were imaged using a Typhoon Imager (GE Healthcare) or a BioRad ChemiDoc. Primary and secondary antibodies are in Supplementary Table 2.

***In vitro* binding assays using recombinant proteins.** Saturating amounts of clarified bacterial lysates containing GST and GST-S fusions were added to Glutathione Sepharose beads that were previously washed with lysis buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 5 mM 2-mercaptoethanol, 1% Triton X-100) and incubated at 4ºC for 1 hour on a tube roller. Beads were washed once with lysis buffer, once with lysis buffer supplemented with 500 mM NaCl, once again with lysis buffer, and incubated with

clarified bacterial lysates containing recombinant WIPI3-6xHis, moesin(1-300)-6xHis, SNX27-6xHis, or 6xHis-VPS26 for 2 hours at 4ºC on a rotator. Beads were washed 3 times with lysis buffer and eluted with 25 mM reduced glutathione made in lysis buffer, supplemented with a 4x solution of NuPage LDS sample buffer containing 100 mM DTT and loaded on SDS PAGE gels (Invitrogen, XP04205) and analysed by Coomassie protein stain (Abcam, ab119211) or by immunoblot using an anti-6xHis HRP-conjugated antibody.

GFP co-immunoprecipitation. 293T cells were seeded in 150 mm dishes in culture medium in a humidified incubator at 37°C with 5% CO_2. Once cells reached a density of approximately 50-70%, they were transfected with 25 µg per plate of plasmid DNA encoding GFP or GFP-CD86-S chimeric fusions using PEI. 48 hours after transfection, cells were mechanically resuspended in culture medium and EDTA solution and centrifuged at 300 x *g* for 5 minutes at 4°C. Cells were washed once with ice cold PBS by resuspension and centrifugation and the pellet was subsequently resuspended in 1 ml of ice cold lysis buffer containing 50 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 1 mM PMSF (Sigma) and 1x cOmplete EDTA-free protease inhibitor cocktail. Cells were subject to sonication using a Misonix 300 water sonicator for 1 minute; 10 seconds on, 10 seconds off at amplitude 5. Cells were incubated at 4°C with agitation for 10 minutes prior to the clarification of the lysate at 16,100 x *g* for 10 minutes at 4°C. For the sample containing the ss-GFP-CD86-S-HA chimeric fusion, 0.6 µg of anti-HA antibody (Roche) was added to the lysate. A fraction of the input lysate was mixed with 1x NuPage LDS sample buffer, 50 mM TCEP pH 7.0. 100 µl of GFP-Trap magnetic agarose bead slurry (Chromotek) per sample was equilibrated by washing once in ice cold lysis buffer using a magnetic rack. The equilibrated beads were resuspended in lysate and incubated at 4°C with agitation for 2 hours. The sample was washed 5 times in lysis buffer using a magnetic rack and eluted in 2x LDS sample buffer with 50 mM TCEP pH 7.0 by boiling at 90°C for 3 minutes.

Comparison of internal and external levels of S by flow cytometry. U2OS cells were seeded at a density of $2x10^4$ cells/cm^2 in T75 flasks in culture medium in a humidified incubator at 37°C with 5% CO_2. 24 hours after seeding, cells were transfected with 15 µg of plasmid DNA encoding different N-terminally HA-tagged S cytoplasmic tail mutants using PEI. 24 hours after transfection, cells were washed once in EDTA solution and dissociated from the flask in accutase (Sigma) for 2 minutes at 37°C. Cells were washed once in ice cold FACS buffer (2% FCS in PBS) by resuspension and centrifugation at 300

x *g* for 5 minutes at 4°C. The supernatant was removed and approximately 10^6 cells were resuspended in FACS buffer containing an anti-HA AF488 conjugate (1:1000, BioLegend, 901509) and an eFluor 780 fixable viability dye (Thermo Fischer Scientific, 65-0865-14). Cells were incubated on ice, in darkness for 30 minutes. Cells were washed 3 times in FACS buffer and incubated in Cyto-Fast Fix/Perm Buffer (BioLegend, 426803) for 20 minutes at room temperature. Cells were washed once in Cyto-Fast Wash Buffer and incubated in Cyto-Fast Wash Buffer containing an anti-HA AF647 conjugate (1:1000, BioLegend, 682404) for 20 minutes at room temperature. Cells were washed twice with Cyto-Fast Wash Buffer and resuspended in FACS buffer. Cells were strained using a 100 µm filter prior to analysis on an LSRII flow cytometer (BD Biosciences). Data was analysed using FlowJo v10 in which singlets were gated according to forward and side scatter profiles, dead cells were excluded using the viability dye and non-transfected cells were excluded based on their low AF488 and 647 signals. Single colour controls were used to conduct compensation.

Assay of endocytosis of S by flow cytometry. Cells were seeded, transfected and resuspended as described above. The antibody uptake assay was based on the use of quenching to distinguish internal from external material[46]. Approximately 10^6 cells were resuspended in complete medium containing an anti-HA AF488 conjugate (1:1000) and incubated at 37°C for 40 minutes. The cells were washed twice with ice cold FACS buffer and incubated with an anti-AF488 antibody (1:67, A-11094, Thermo Fischer Scientific; to quench non-internalised anti-HA AF488 conjugate), an anti-mouse AF647 antibody (1:300, Thermo Fischer Scientific, A31571; to relabel the non-internalised anti-HA conjugate) and an eFluor 780 fixable viability dye. Cells were washed three times in ice cold FACS buffer, fixed in 4% paraformaldehyde (PFA) for 20 minutes and washed a further two times in FACS buffer. Cells were strained and analysed as described above. Internalised anti-HA AF488 conjugate was inaccessible to quenching and therefore any associated AF488 signal was equated to levels of internalised S. Conversely, only anti-HA AF488 conjugate at the cell surface was accessible by the anti-mouse AF647 secondary antibody and therefore any associated AF647 signal was equated to levels of non-internalised S.

Comparison of internal and external levels of S by immunofluorescence. U2OS cells were seeded at a density of $2x10^4$ cells/cm^2 in 6-well plates in culture medium in a humidified incubator at 37°C with 5% CO_2. 24 hours after seeding, cells were transfected with 1-2 µg of plasmid DNA encoding different N-terminally HA-tagged S cytoplasmic tail

mutants using PEI. 24 hours after transfection, cells were washed once in EDTA solution and dissociated from the flask in trypsin for 2 minutes at 37°C and seeded onto coated microscope slides (Hendley-Essex) in culture medium in a humidified incubator at 37°C with 5% CO_2. 24 hours after seeding, cells were washed with PBS and fixed in 4% PFA in PBS for 20 minutes at room temperature. Cells were incubated in an anti-HA AF647 conjugate (1:1000) diluted in 20% FCS in PBS for 20 minutes in order to label HA-tagged S at the surface of cells. Cells were washed three times in PBS and permeabilised in 10% Triton X-100 in PBS for 10 minutes. Cells were subsequently washed twice in PBS and incubated in blocking buffer (20% FCS, 1% Tween-20 in PBS) for 1 hour. Cells were incubated with an anti-calreticulin antibody (1:200, Abcam, ab2907) diluted in blocking buffer for 1 hour in order to counterstain the ER. Cells were washed twice in PBS, incubated in blocking buffer for 10 minutes and washed twice again in PBS. Cells were incubated for 1 hour in darkness with an anti-rabbit AF555 secondary antibody (1:300, Thermo Fischer Scientific, A31572) and an anti-HA AF488 conjugate (1:1000) diluted in blocking buffer with the latter intended to label HA-tagged S associated with intracellular membranes (as well as available external HA-S epitopes). Cells were washed twice in PBS, incubated in blocking buffer for 10 minutes and washed twice again in PBS. Cells were mounted in Vectashield (Vector Laboratories) prior to application of a coverslip which was sealed using nail vanish. Slides were imaged using a Leica TCS SP8 confocal microscope.

S protein uptake assay. U2OS cells were transfected with plasmids encoding N-terminally HA-tagged S and variants, or HA-tagged CPD, and seeded onto microscope slides as described above. 48 hours after transfection, cells were incubated in ice cold media and slides were placed on ice for 15 minutes. An anti-HA antibody (1:300, Roche, 3F10) diluted in ice cold media was added and cells were incubated on ice for 30 minutes. Cells were washed twice with ice cold PBS and subject to a chase in which warm media was added and cells were incubated in a humidified incubator at 37°C with 5% CO_2 for 1 hour. Cells were washed, fixed and blocked with 20% FCS in PBS as described above. Cells were incubated with an anti-rat AF647 secondary antibody (1:300, Abcam, ab150154) diluted in 20% FCS in PBS to label non-internalised anti-HA antibody under non-permeabilising conditions at room temperature for 20 minutes. Cells were washed three times in PBS and permeabilised and blocked as described above. Cells were incubated with an anti-SNX27 antibody (1:300, Abcam, ab77799) in blocking buffer for 1 hour in order to counterstain endosomes. Cells were washed twice in PBS, incubated in blocking buffer for 10 minutes and washed twice again in PBS. Cells were

incubated for 1 hour in darkness with an anti-mouse AF555 secondary antibody (1:300, Thermo Fischer Scientific, A31571) and an anti-rat AF488 secondary antibody (1:300, Thermo Fischer Scientific, A21208) diluted in blocking buffer with the latter intended to stain internalised anti-HA antibody. Cells were washed, mounted and imaged as described above.

Cell-cell fusion assay. Acceptor and donor cells were seeded at 70% confluency in a 24-well plate 16 hours prior transfection. Donor 293T cells were co-transfected with 1.5 µg of plasmids encoding different untagged S mutants and 0.6 µg of pmCherry-N1 using 6 µL of Fugene 6 following the manufacturer's instructions (Promega). Acceptor Vero cells were treated with CellTracker™ Green CMFDA (5-chloromethylfluorescein diacetate, Thermo Scientific) for 30 minutes. Donor cells were then detached 5 hours post transfection, mixed together with the green-labelled acceptor cells and plated in a 12-well plate. Cell-cell fusion events was measured using an IncuCyte (Sartorious) and determined as the proportion of merged area to green area over time. Data were analysed using IncuCyte software and plotted with Prism 8.

References

1. Hoffmann, M. *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and Is blocked by a clinically proven protease inhibitor. *Cell* **181,** 271–280 (2020).
2. Bao, L. *et al.* The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature* **583,** 830–833 (2020).
3. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581,** 221–224 (2020).
4. Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367,** 1444–1448 (2020).
5. Walls, A. C. *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181,** 281–292 (2020).
6. Wang, Y.-T. *et al.* Spiking pandemic potential: structural and immunological aspects of SARS-CoV-2. *Trends in Microbiology* **28,** 605–618 (2020).
7. Youn, S., Collisson, E. W. & Machamer, C. E. Contribution of trafficking signals in the cytoplasmic tail of the infectious bronchitis virus spike protein to virus infection. *J Virol* **79,** 13209–13217 (2005).

8. Hou, Y., Meulia, T., Gao, X., Saif, L. J. & Wang, Q. Deletion of both the tyrosine-based endocytosis signal and the endoplasmic reticulum retrieval signal in the cytoplasmic tail of spike protein attenuates porcine epidemic diarrhea virus in pigs. *J Virol* **93,** e01758–18 (2019).
9. Buchrieser, J. *et al.* Syncytia formation by SARS-CoV-2 infected cells. *bioRxiv* 2020.07.14.202028 (2020).
10. Zhu, N. *et al.* Morphogenesis and cytopathic effect of SARS-CoV-2 infection in human airway epithelial cells. *Nat Commun* **11,** 3910–8 (2020).
11. Tian, S. *et al.* Pulmonary pathology of early-phase 2019 novel coronavirus (COVID-19) pneumonia in two patients with lung cancer. *J Thorac Oncol* **15,** 700–704 (2020).
12. J Alsaadi, E. A. & Jones, I. M. Membrane binding proteins of coronaviruses. *Future Virol* **14,** 275–286 (2019).
13. Ujike, M. & Taguchi, F. Incorporation of spike and membrane glycoproteins into coronavirus virions. *Viruses* **7,** 1700–1725 (2015).
14. Li, F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol* **3,** 237–261 (2016).
15. Forni, D., Cagliani, R., Clerici, M. & Sironi, M. Molecular evolution of human coronavirus genomes. *Trends in Microbiology* **25,** 35–48 (2017).
16. Neuman, B. W. & Buchmeier, M. J. Supramolecular architecture of the coronavirus particle. *Adv. Virus Res.* **96,** 1–27 (2016).
17. Nal, B. *et al.* Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. *J. Gen. Virol.* **86,** 1423–1434 (2005).
18. Klein, S. *et al.* SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. bioRxiv 2020.06.23.167064 (2020).
19. Tooze, J., Tooze, S. A. & Fuller, S. D. Sorting of progeny coronavirus from condensed secretory proteins at the exit from the trans-Golgi network of AtT20 cells. *J Cell Biol* **105,** 1215–1226 (1987).
20. Tseng, C.-T. K. *et al.* Apical entry and release of severe acute respiratory syndrome-associated coronavirus in polarized Calu-3 lung epithelial cells. *J Virol* **79,** 9470–9479 (2005).
21. Millet, J. K. *et al.* Ezrin interacts with the SARS coronavirus Spike protein and restrains infection at the entry stage. *PLoS ONE* **7,** e49566 (2012).
22. McBride, C. E., Li, J. & Machamer, C. E. The cytoplasmic tail of the severe acute respiratory syndrome coronavirus spike protein contains a novel endoplasmic

reticulum retrieval signal that binds COPI and promotes interaction with membrane protein. *J Virol* **81,** 2418–2428 (2007).

23. Gallon, M. *et al.* A unique PDZ domain and arrestin-like fold interaction reveals mechanistic details of endocytic recycling by SNX27-retromer. *Proc Natl Acad Sci U S A* **111,** E3604–13 (2014).
24. Bakula, D. *et al.* WIPI3 and WIPI4 β-propellers are scaffolds for LKB1-AMPK-TSC signalling circuits in the control of autophagy. *Nat Commun* **8,** 15637 (2017).
25. McBride, C. E. & Machamer, C. E. Palmitoylation of SARS-CoV S protein is necessary for partitioning into detergent-resistant membranes and cell-cell fusion but not interaction with M protein. *Virology* **405,** 139–148 (2010).
26. Petit, C. M. *et al.* Palmitoylation of the cysteine-rich endodomain of the SARS-coronavirus spike glycoprotein is important for spike-mediated cell fusion. *Virology* **360,** 264–274 (2007).
27. Votsmeier, C. & Gallwitz, D. An acidic sequence of a putative yeast Golgi membrane protein binds COPII and facilitates ER export. *EMBO J* **20,** 6742–6750 (2001).
28. Gomez-Navarro, N. & Miller, E. Protein sorting at the ER-Golgi interface. *J Cell Biol* **215,** 769–778 (2016).
29. Jackson, L. P. *et al.* Molecular basis for recognition of dilysine trafficking motifs by COPI. *Dev Cell* **23,** 1255–1262 (2012).
30. Teasdale, R. D. & Jackson, M. R. Signal-mediated sorting of membrane proteins between the endoplasmic reticulum and the Golgi apparatus. *Annu Rev Cell Dev Biol* **12,** 27–54 (1996).
31. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583,** 459–468 (2020).
32. Stukalov, A. *et al.* Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. **193,** 3080–38 (2020).
33. Papa, G. *et al.* Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. *bioRxiv* 2020.08.13.243303 (2020).
34. Ma, W. & Goldberg, J. Rules for the recognition of dilysine retrieval motifs by coatomer. *EMBO J* **32,** 926–937 (2013).
35. Schwegmann-Weßels, C. *et al.* A novel sorting signal for intracellular localization is present in the S protein of a porcine coronavirus but absent from severe acute respiratory syndrome-associated coronavirus. *J Biol Chem* **279,** 43661–43666 (2004).

36. Winter, C., Schwegmann-Weßels, C., Neumann, U. & Herrler, G. The spike protein of infectious bronchitis virus is retained intracellularly by a tyrosine motif. *J Virol* **82,** 2765–2771 (2008).
37. Michie, K. A., Bermeister, A., Robertson, N. O., Goodchild, S. C. & Curmi, P. M. G. Two Sides of the Coin: Ezrin/Radixin/Moesin and Merlin Control Membrane Structure and Contact Inhibition. *Int J Mol Sci* **20,** (2019).
38. Huang, Y., Yang, Z.-Y., Kong, W.-P. & Nabel, G. J. Generation of synthetic severe acute respiratory syndrome coronavirus pseudoparticles: implications for assembly and vaccine production. *J Virol* **78,** 12557–12565 (2004).
39. Bussani, R. *et al.* Persistence of viral RNA, widespread thrombosis and abnormal cellular syncytia are hallmarks of COVID-19 lung pathology. *bioRxiv* 2020.06.22.20136358 (2020).
40. Sattentau, Q. Avoiding the void: cell-to-cell spread of human viruses. *Nat Rev Micro* **6,** 815–826 (2008).
41. Cifuentes-Muñoz, N., Dutch, R. E. & Cattaneo, R. Direct cell-to-cell transmission of respiratory viruses: The fast lanes. *PLoS Pathog* **14,** e1007015 (2018).
42. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47,** D442–D450 (2019).
43. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26,** 1367–1372 (2008).
44. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10,** 1794–1805 (2011).
45. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Meth* **13,** 731–740 (2016).
46. Kozik, P., Francis, R. W., Seaman, M. N. J. & Robinson, M. S. A screen for endocytic motifs. *Traffic* **11,** 843–855 (2010).

Figure Legends

Fig. 1. Proteomic analysis of the binding partners of the cytoplasmic tail of SARS-CoV-2 S protein. a, Mass spectrometry analysis from affinity chromatography of 293T cell lysates using GST-S tail (1237-1273). The plot compares the average spectral counts from three independent replicates of GST-S(1237-1273) versus the negative control (GST). Values are in Supplementary Table 1. **b,c,** Immunoblots of eluates from the indicated GST-tagged S tails as prepared in (**a**). Coomassie blue-stained gels show the GST-tail fusions with adjacent residues of the tail of S (residues 1237-1273) mutated to alanine (**b**). The blots shown are representative from two independent experiments, and the input lysate represents 1/100 of the material applied to the GST fusions. **c**, Validation of the interactions using two distinct halves of the tail of S: the membrane proximal half (residues 1237-1254) and the distal half (residues 1255-1273). The blots shown are representative from three independent experiments. **d**, Schematic of the transmembrane and cytoplasmic domain of the SARS-CoV-2 S protein. The residues that are critical for binding to the different cytoplasmic factors are indicated.

Fig. 2. The tail of SARS-CoV-2 S protein binds directly to moesin, WIPI3, SNX27, and to VPS26 via SNX27. **a**, Coomassie-stained gel and immunoblot against 6xHis to test the binding of the FERM domain of moesin (residues 1-300, expressed in bacterial lysate) to GST fusions to the tail of S. Experiment repeated twice, input 1/40 that applied to beads. **b**, As in (**a**), beads coated with the indicated GST-fusions were incubated with clarified bacterial lysates from cells overexpressing WIPI3-6xHis. Experiment repeated twice, input 1/7 that applied to beads. **c,** Coomassie gel showing that the residues 1238ThrSerCys1240 of the tail of S are required for binding to SNX27. As in (**a)** except that the lysate was from cells overexpressing SNX27-6xHis. Experiment repeated three times, input 1/40 that applied to beads. **d**, Coomassie gel showing that VPS26 requires SNX27 to bind to the tail of S protein. Same experiment as in (**a**) except that beads coated with the GST-fusions were incubated with bacterial lysate from cells overexpressing either SNX27-6xHis, 6xHis-VPS26 or both. Experiment repeated twice.

Fig. 3. The cytoplasmic tail of the SARS-CoV-2 S protein harbours di-acidic ER export motifs and a suboptimal ER retrieval motif. **a**, C-terminal sequences of the D/E to A mutant of S indicating the putative di-acidic ER export motifs mutated to alanine. **b**, Micrographs of U2OS cells transiently expressing N-terminally HA-tagged wild-type S and the D/E to A mutant. Cell surface S was initially stained using an anti-HA Alexa Fluor

AF647 under non-permeabilising conditions. Cells were subsequently permeabilised and stained with anti-HA AF488 to show the internal S. Scale bars 10 μm. Experiment repeated twice. **c**, Quantification of (**b**) by flow cytometry. Displayed are overlaid histograms representing the ratio of extracellular S (AF488 signal) to that of intracellular S (AF647 signal) for wild-type S and the D/E to A mutant. Histograms normalised to the mode value and represent >10,000 events, N = 3. **d**, Schematic of the cytoplasmic tail of S protein showing three different mutations made in the COPI-binding site. **e**, Immunoblots of GFP-trap co-immunoprecipitation experiments from 293T cells overexpressing GFP-CD86/S cytoplasmic tail chimeras. The immunoprecipitation of the wild-type-HA sample was done in the presence of an anti-HA antibody. α-tubulin (loading control). N = 2. **f**, Beads coated with the indicated recombinant S tail GST-fusions were incubated with clarified detergent lysate from 293T cells and eluates immunoblotted for β-COP. The H1271K mutation has increased affinity for the COPI coat. Representative blot from two independent experiments, input 1/50 of that applied to beads. **g**, Flow cytometry as in (**c**) to quantify the phenomenon observed in (**f**), N = 3. **h**, Micrographs as in (**b**) but with mutants which bind differentially to COPI. Scale bars 10 μm. Experiment repeated twice.

Fig. 4. SARS-CoV-2 S leakage to the plasma membrane facilitates cell-cell fusion and the formation of syncytia. **a**, Comparison of the cytoplasmic tails of S proteins from the four genera of Coronaviridae. Binding motifs for trafficking proteins are indicated: COPI (green), COPII (blue), and clathrin adaptors (red), along with cysteine residues (brown). Retention of S in the ER reduces its capacity to form syncytia while increased leakage to the plasma membrane increases it. **b**, Representative micrographs showing 293T cells coexpressing mCherry and different full-length S variants fusing with acceptor Vero cells stained with CellTracker Green. Cell fusion is monitored by overlap between the green and red signal. Scale bars: 200 μm. **b**, Quantification of cell-cell fusion as percentage of field showing green and red overlap area. N = 2, error bars SEM. **c**, Immunoblots from 293T cells expressing the indicated full-length S variants, with the H1271K mutation restricting the protein in the early Golgi and thus away from the site of S1/S2 cleavage by furin.

Extended Data Fig. 1. Affinity chromatography of cell lysate with the cytoplasmic tail of the SARS-CoV-2 S protein. a, Volcano plot comparing the spectral intensities from proteins bound to GST-S(1237-1273) or GST alone, using data from three independent experiments. **b**, Plot comparing the spectral counts of proteins found by

mass spectrometry analysis of interactors of membrane proximal GST-S(1237-1254 versus the membrane distal GST-S(1255-1273). The spectral counts are means from two biological repeats for GST-S(1237-1254) and three repeats for GST-S(1255-1273). All values in Supplementary Table 1.

Extended Data Fig. 2. Analysis of the binding of WIPI3 and SNX27/retromer to the cytoplasmic tail of S protein. a, The indicated GST-fusions expressed in E. coli, purified with glutathione Sepharose beads, and incubated with a 293T cell lysate followed by analysis of the bound proteins by immunoblot with the indicated antibodies. Coomassie-stained gel shows the GST-tail fusions. For WIPI3-HA, cells were transfected with a plasmid expressing the tagged protein and the blot probed for the HA tag. Representative blots from three independent experiments, input 1/50 of the lysate applied to the beads. **b**, Mapping of residues required for binding to SNX27/retromer using the indicated GST-fusions as in (a) followed by probing with an anti-SNX27 antibody conjugated to HRP. Representative blots from three independent experiments, input 1/50 of lysate applied to beads.

Extended Data Fig. 3. The SARS-CoV-2 S protein is not rapidly internalised from the cell surface. a, Micrographs of U2OS cells transiently expressing variants of S or carboxypeptidase D (CPD) (all with N-terminal HA tags), and subjected to an antibody uptake assay. Cells were incubated with an anti-HA antibody in complete medium for 30 minutes on ice, unbound antibody washed off and cells chased for one hour in complete medium at 37ºC. Cells were fixed and surface S was labelled using an AF647-conjugated secondary antibody under non-permeabilising conditions. Cells were then permeabilised and stained with an AF488-conjugated secondary antibody to label both surface and internalised S. Scale bars:10 µm. **b**, Flow cytometry analysis comparing HA-tagged S and CPD in an antibody uptake assay. Cells were stained with an anti-HA AF488 conjugate for 30 minutes on ice, unbound antibody was washed off and cells chased for 40-minutes at 37ºC. Cells were fixed and non-internalised anti-HA AF488 conjugated antibody was simultaneously quenched and relabelled with AF647 by incubation with an anti-AF488 antibody and an AF647-conjugated secondary antibody respectively under non-permeabilising conditions. Histograms represent the ratio of internalised S (AF488 signal) to that of non-internalised S (AF647 signal), and are normalised to the mode value and represent 5000 - 10,000 events, N = 2.

Figure 1

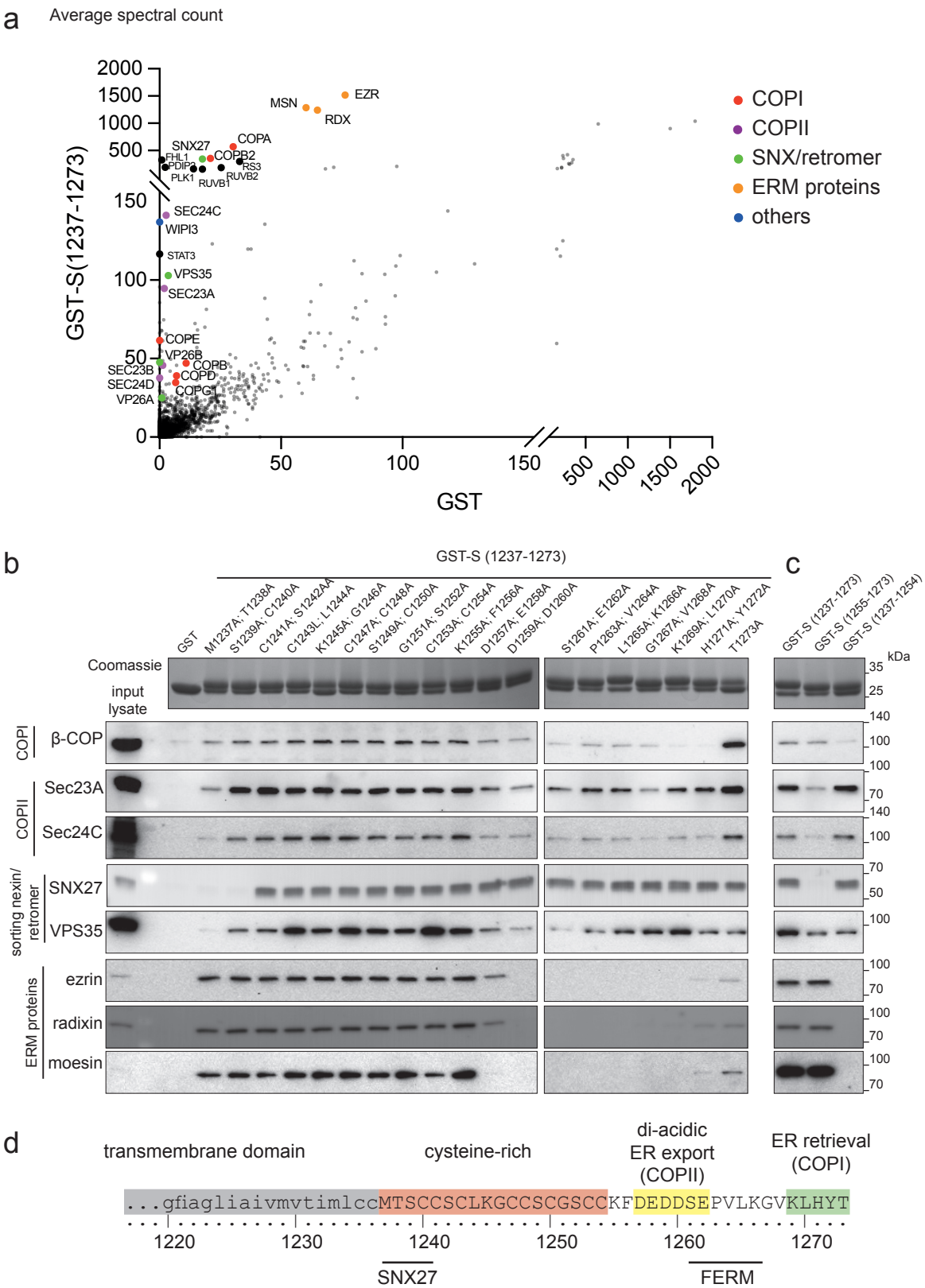

a
Average spectral count
GST-S(1237-1273)
GST
COPI
COPII
SNX/retromer
ERM proteins
others
EZR
MSN
RDX
COPA
SNX27
COPB2
FHL1
PDIP
PLK1
RUVB1
RS3
RUVB2
SEC24C
WIPI3
STAT3
VPS35
SEC23A
COPE
VP26B
COPB
SEC23B
COPD
SEC24D
COPG1
VP26A
b
GST-S (1237-1273)
GST
M1237A; T1238A
S1239A; C1240A
C1241A; S1242AA
C1243L; L1244A
K1245A; G1246A
C1247A; C1248A
S1249A; C1250A
G1251A; S1252A
C1253A; C1254A
K1255A; F1256A
D1257A; E1258A
D1259A; D1260A
S1261A; E1262A
P1263A; V1264A
L1265A; K1266A
G1267A; V1268A
K1269A; L1270A
H1271A; Y1272A
T1273A
Coomassie
input lysate
COPI
β-COP
COPII
Sec23A
Sec24C
sorting nexin/ retromer
SNX27
VPS35
ERM proteins
ezrin
radixin
moesin
c
GST-S (1237-1273)
GST-S (1255-1273)
GST-S (1237-1254)
kDa
d
transmembrane domain
cysteine-rich
di-acidic ER export (COPII)
ER retrieval (COPI)
...gfiagliaivmvtimlccMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT
1220
1230
1240
1250
1260
1270
SNX27
FERM

Figure 2

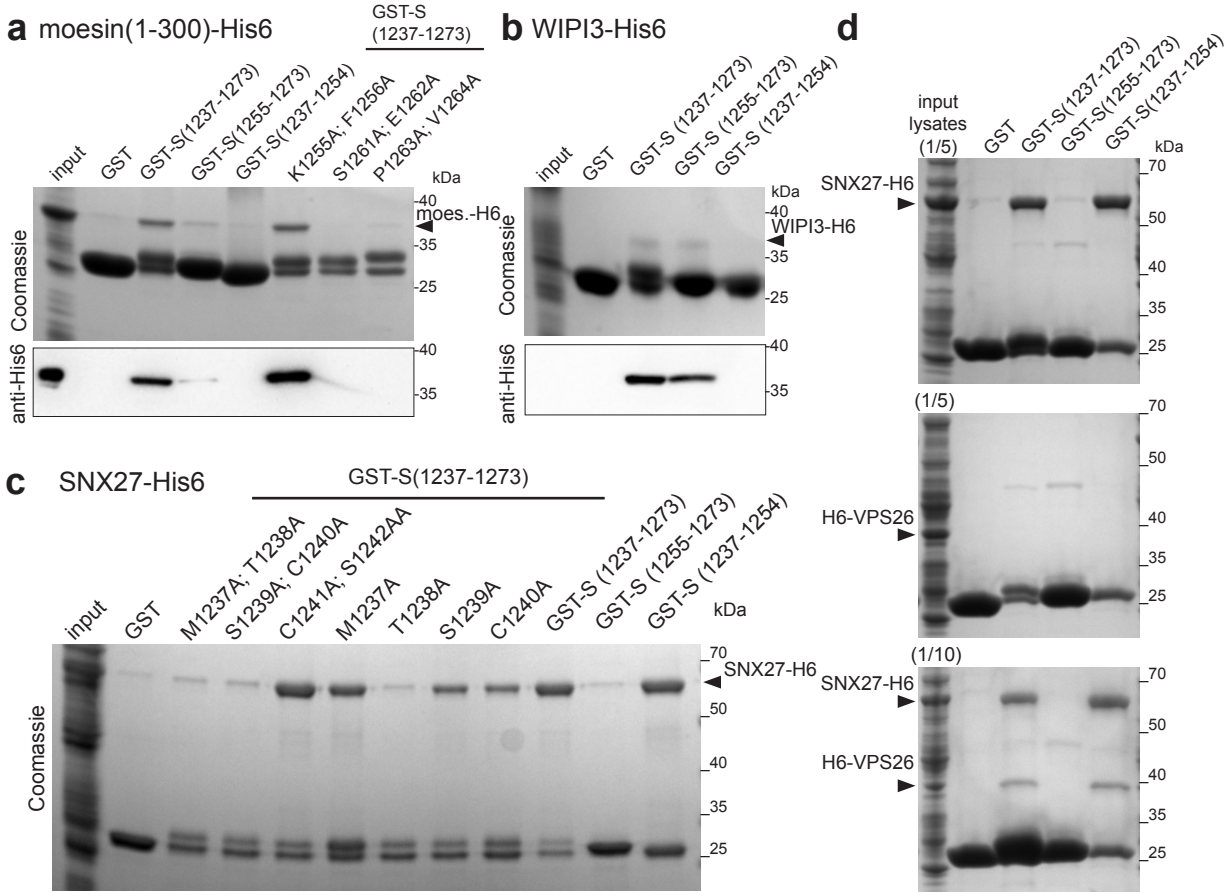

a moesin(1-300)-His6
GST-S (1237-1273)
input
GST
GST-S(1237-1273)
GST-S(1255-1273)
GST-S(1237-1254)
K1255A; F1256A
S1261A; E1262A
P1263A; V1264A
kDa
moes.-H6
Coomassie
anti-His6
b WIPI3-His6
GST-S (1237-1273)
GST-S (1255-1273)
GST-S (1237-1254)
WIPI3-H6
c SNX27-His6
GST-S(1237-1273)
M1237A; T1238A
S1239A; C1240A
C1241A; S1242AA
M1237A
T1238A
S1239A
C1240A
SNX27-H6
d
input lysates (1/5)
(1/5)
(1/10)
H6-VPS26

Figure 3

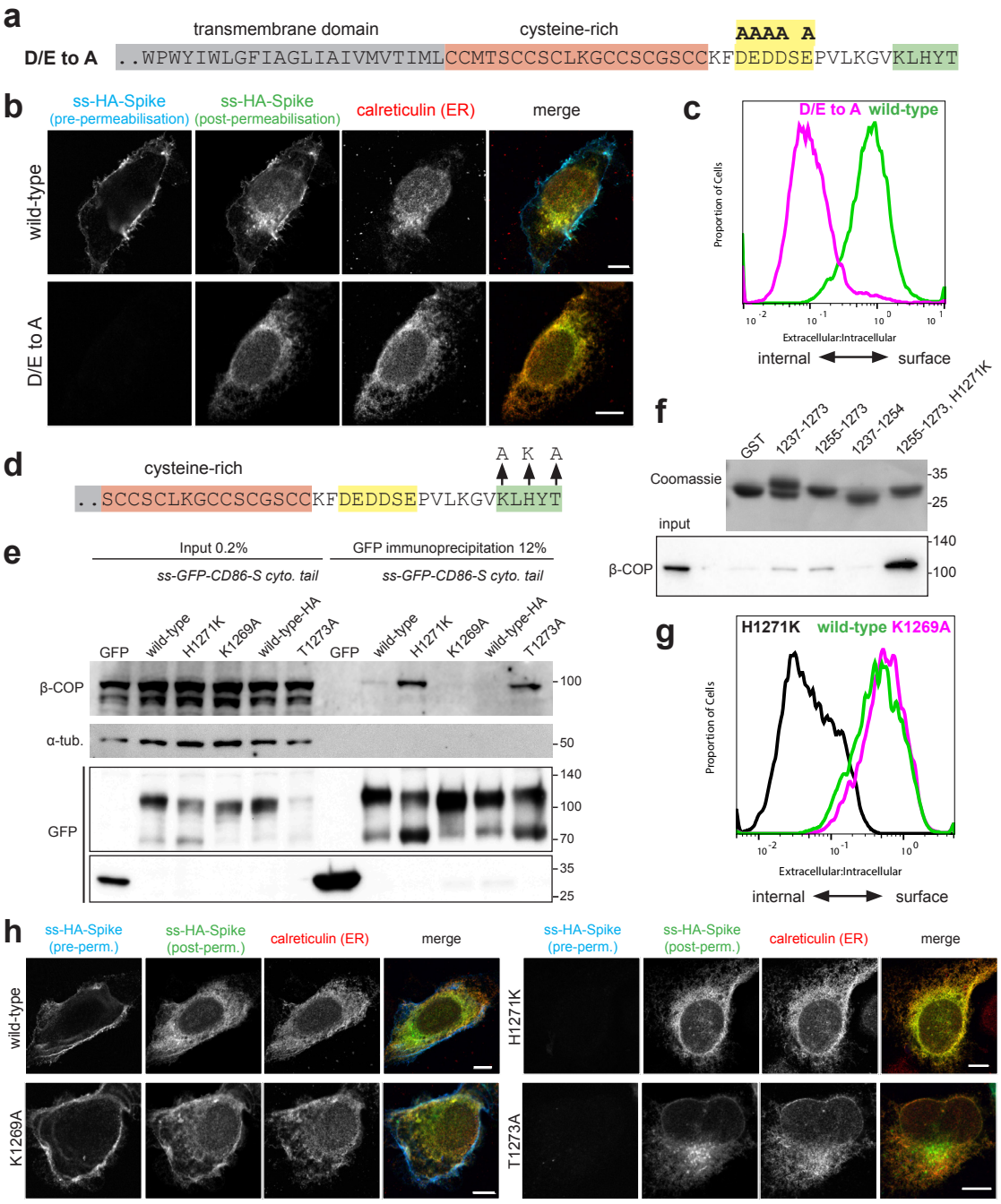
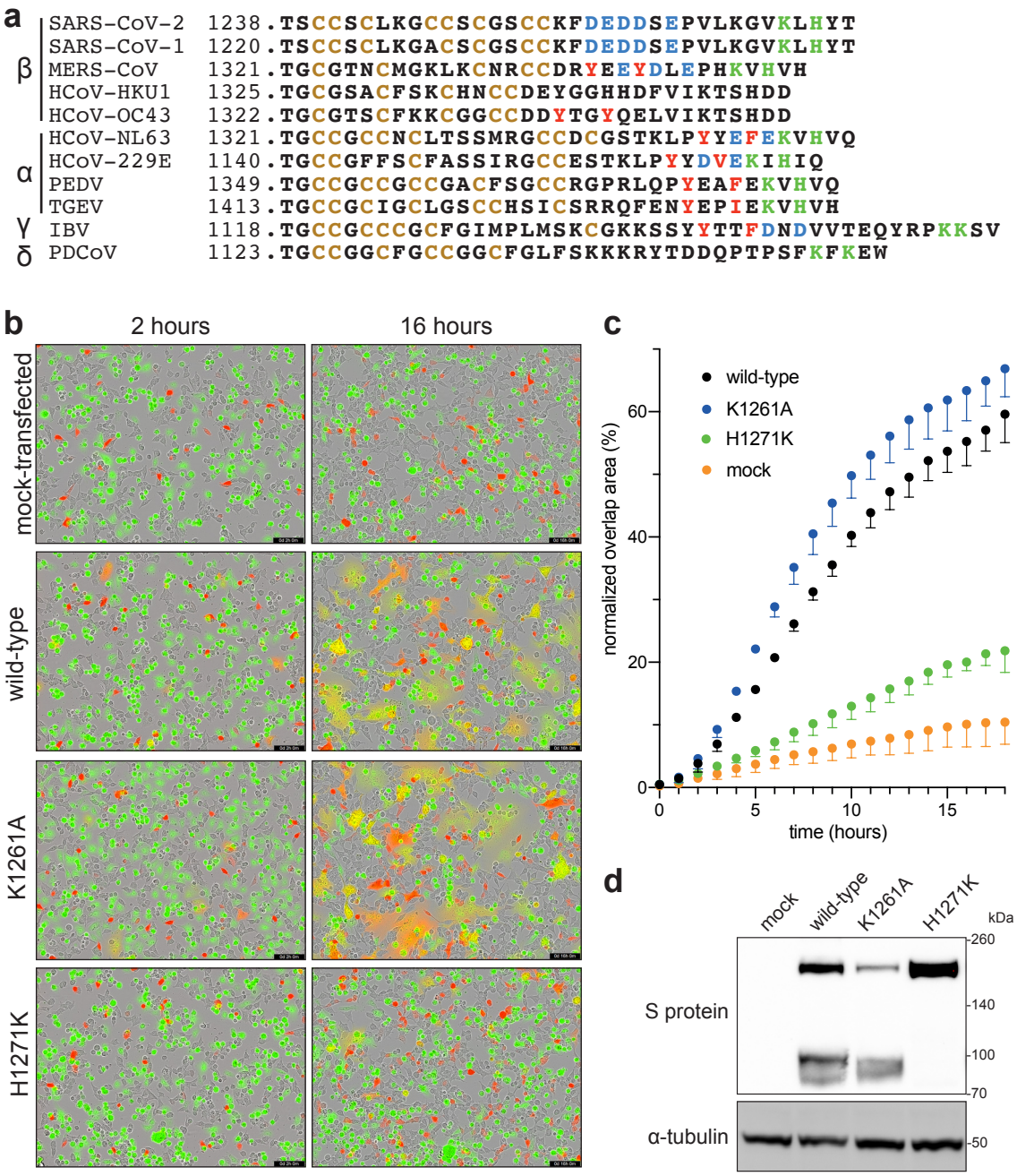
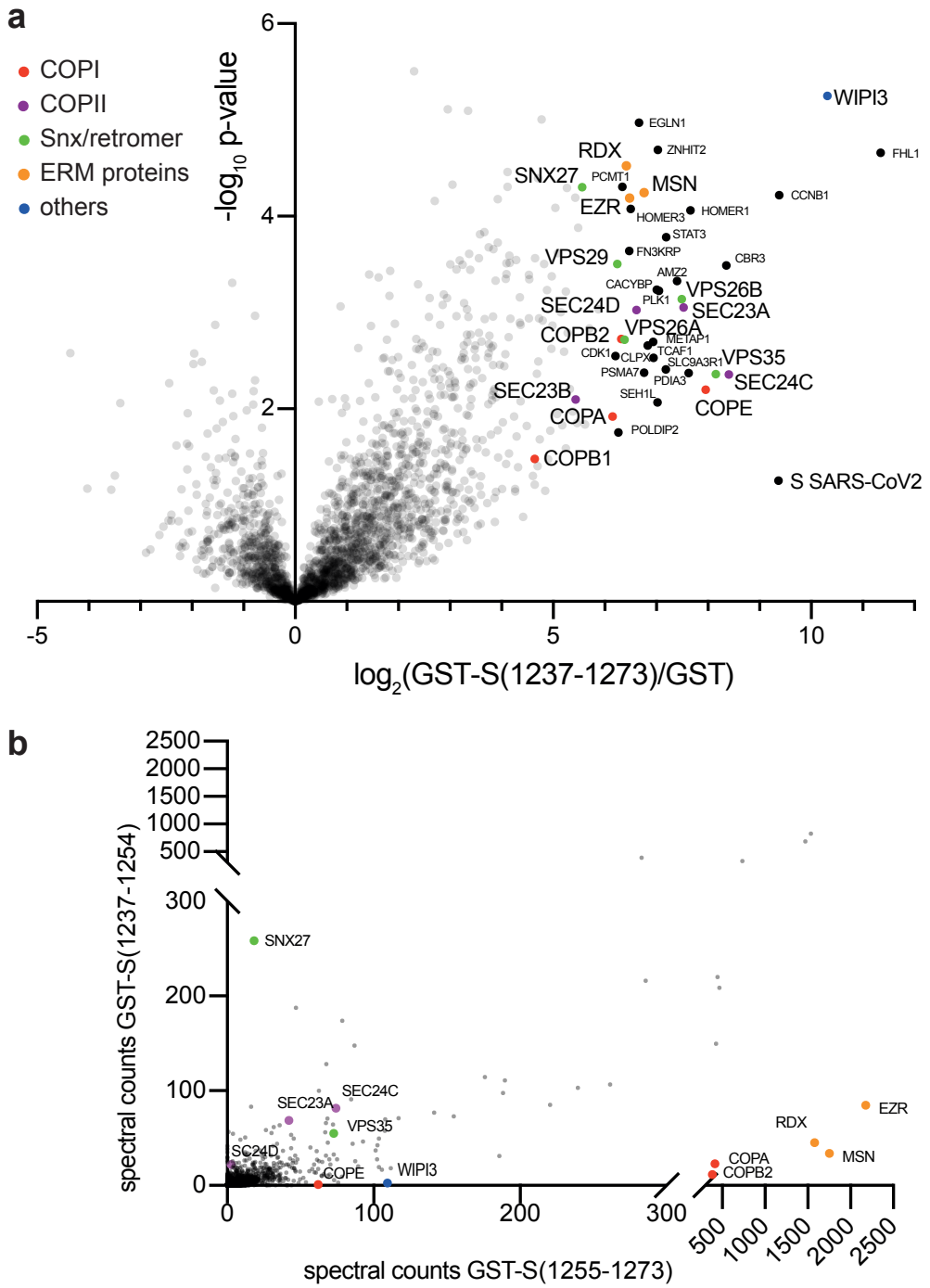
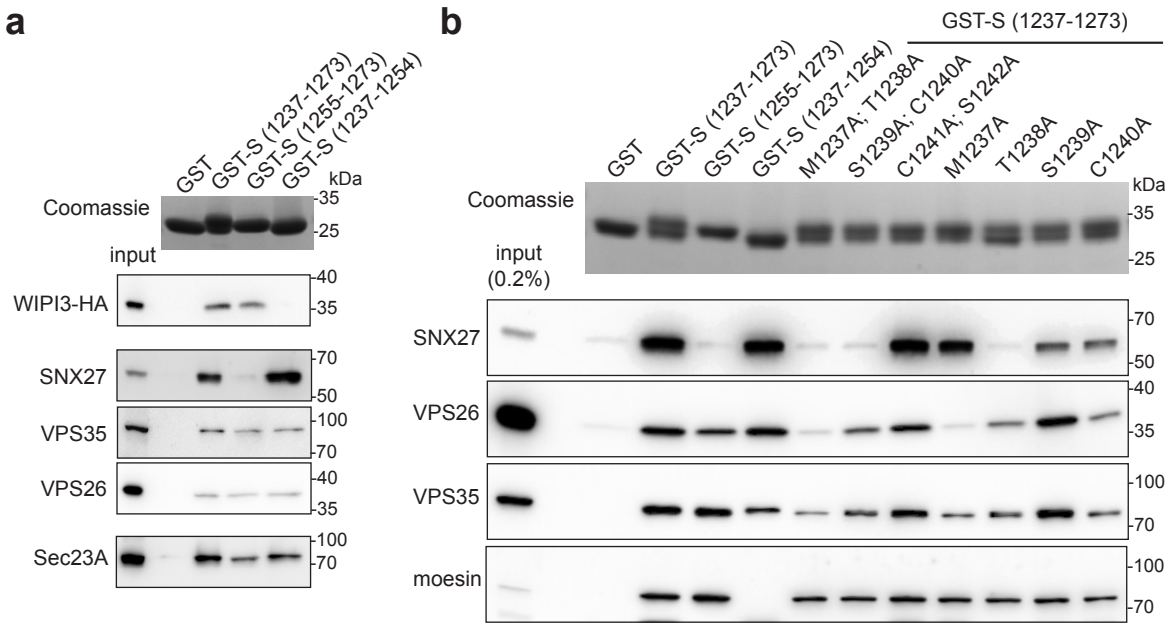

a
transmembrane domain
cysteine-rich
AAAA A
D/E to A
..WPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT
b
ss-HA-Spike (pre-permeabilisation)
ss-HA-Spike (post-permeabilisation)
calreticulin (ER)
merge
wild-type
D/E to A
c
D/E to A
wild-type
Proportion of Cells
Extracellular:Intracellular
internal
surface
d
cysteine-rich
A K A
..SCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT
e
Input 0.2%
GFP immunoprecipitation 12%
ss-GFP-CD86-S cyto. tail
ss-GFP-CD86-S cyto. tail
GFP
wild-type
H1271K
K1269A
wild-type-HA
T1273A
β-COP
α-tub.
GFP
100
50
140
100
70
35
25
f
GST
1237-1273
1255-1273
1237-1254
1255-1273, H1271K
Coomassie
input
β-COP
35
25
140
100
g
H1271K
wild-type
K1269A
Proportion of Cells
Extracellular:Intracellular
internal
surface
h
ss-HA-Spike (pre-perm.)
ss-HA-Spike (post-perm.)
calreticulin (ER)
merge
wild-type
H1271K
K1269A
T1273A

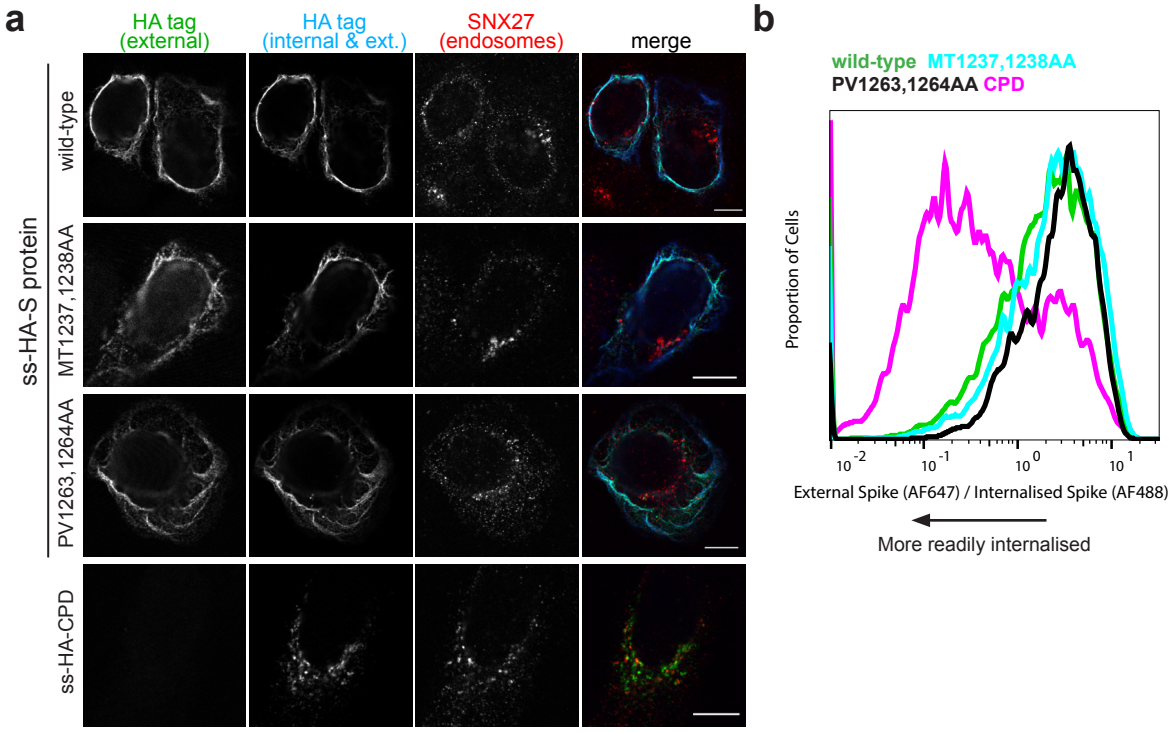
Figure 4


a
β
α
γ
δ
SARS-CoV-2 1238.TSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT
SARS-CoV-1 1220.TSCCSCLKGACSCGSCCKFDEDDSEPVLKGVKLHYT
MERS-CoV 1321.TGCGTNCMGKLKCNRCCDRYEEYDLEPHKVHVH
HCoV-HKU1 1325.TGCGSACFSKCHNCCDEYGGHHDFVIKTSHDD
HCoV-OC43 1322.TGCGTSCFKKCGGCCDDYTGYQELVIKTSHDD
HCoV-NL63 1321.TGCCGCCNCLTSSMRGCCDCGSTKLPYYEFEKVHVQ
HCoV-229E 1140.TGCCGFFSCFASSIRGCCESTKLPYYDVEKIHIQ
PEDV 1349.TGCCGCCGCCGACFSGCCRGPRLQPYEAFEKVHVQ
TGEV 1413.TGCCGCIGCLGSCCHSICSRRQFENYEPIEKVHVH
IBV 1118.TGCCGCCCGCFGIMPLMSKCGKKSSYYTTFDNDVVTEQYRPKKSV
PDCoV 1123.TGCCGGCFGCCGGCFGLFSKKKRYTDDQPTPSFKFKEW
b
2 hours
16 hours
mock-transfected
wild-type
K1261A
H1271K
c
wild-type
K1261A
H1271K
mock
normalized overlap area (%)
0
20
40
60
0
5
10
15
time (hours)
d
mock
wild-type
K1261A
H1271K
kDa
260
140
100
70
50
S protein
α-tubulin

Extended Data Fig. 1




a
GST
GST-S (1237-1273)
GST-S (1255-1273)
GST-S (1237-1254)
kDa
Coomassie
35
25
input
WIPI3-HA
40
35
SNX27
70
50
VPS35
100
70
VPS26
40
35
Sec23A
100
70
b
GST-S (1237-1273)
GST
GST-S (1237-1273)
GST-S (1255-1273)
GST-S (1237-1254)
M1237A; T1238A
S1239A; C1240A
C1241A; S1242A
M1237A
T1238A
S1239A
C1240A
kDa
Coomassie
35
25
input
(0.2%)
SNX27
70
50
VPS26
40
35
VPS35
100
70
moesin
100
70


a
HA tag (external)
HA tag (internal & ext.)
SNX27 (endosomes)
merge
ss-HA-S protein
wild-type
MT1237,1238AA
PV1263,1264AA
ss-HA-CPD
b
wild-type MT1237,1238AA
PV1263,1264AA CPD
Proportion of Cells
10 -2
10 -1
10 0
10 1
External Spike (AF647) / Internalised Spike (AF488)
More readily internalised