

# A call for an independent inquiry into the origin of the SARS-CoV-2 virus

Neil L. Harrison<sup>a,b,1</sup> and Jeffrey D. Sachs<sup>c,1</sup>



Since the identification of the SARS-CoV-2 in Wuhan, China, in January 2020 (1), the origin of the virus has been a topic of intense scientific debate and public speculation. The two main hypotheses are that the virus emerged from human exposure to an infected animal ["zoonosis" (2)] or that it emerged in a research-related incident (3). The investigation into the origin of the virus has been made difficult by the lack of key evidence from the earliest days of the outbreak—there's no doubt that greater transparency on the part of Chinese authorities would be enormously helpful. Nevertheless, we argue here that there is much important information that can be gleaned from US-based research institutions, information not yet made available for independent, transparent, and scientific scrutiny.

The data available within the United States would explicitly include, but are not limited to, viral sequences gathered and held as part of the PREDICT project and other funded programs, as well as sequencing data and laboratory notebooks from US laboratories. We call on US government scientific agencies, most notably the NIH, to support a full, independent, and transparent investigation of the origins of SARS-CoV-2. This should take place, for example, within a tightly focused science-based bipartisan Congressional inquiry with full investigative powers, which would be able to ask important questions—but avoid misguided witch-hunts governed more by politics than by science.

**When it comes to deciphering the origins of COVID-19, much important information can be gleaned from US-based research institutions—information that has yet to be made available for independent, transparent, and scientific scrutiny. Image credit: Dave Cutler (artist).**

Author contributions: N.L.H. and J.D.S. wrote the paper.

The authors declare no competing interest.

Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and have not been endorsed by the National Academy of Sciences.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: nh2298@columbia.edu or sachs@ei.columbia.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2202769119/-/DCSupplemental>.

Published May 19, 2022.

## Essential US Investigations

The US intelligence community (IC) was tasked, in 2021 by President Joe Biden (4), with investigating the origin of the virus. In their summary public statement, the IC writes that “all agencies assess that two hypotheses are plausible: natural exposure to an infected animal and a laboratory-associated incident” (4). The IC further writes that “China’s cooperation most likely would be needed to reach a conclusive assessment of the origins of COVID-19 [coronavirus disease 2019].” Of course, such cooperation is highly warranted and should be pursued by the US Government and the US scientific community. Yet, as outlined below, much could be learned by investigating US-supported and US-based work that was underway in collaboration with Wuhan-based institutions, including the Wuhan Institute of Virology (WIV), China. It is still not clear whether the IC investigated these US-supported and US-based activities. If it did, it has yet to make any of its findings available to the US scientific community for independent and transparent analysis and assessment. If, on the other hand, the IC did not investigate these US-supported and US-based activities, then it has fallen far short of conducting a comprehensive investigation.

This lack of an independent and transparent US-based scientific investigation has had four highly adverse consequences. First, public trust in the ability of US scientific institutions to govern the activities of US science in a responsible manner has been shaken. Second, the investigation of the origin of SARS-CoV-2 has become politicized within the US Congress (5); as a result, the inception of an independent and transparent investigation has been obstructed and delayed. Third, US researchers with deep knowledge of the possibilities of a laboratory-associated incident have not been enabled to share their expertise effectively. Fourth, the failure of NIH, one of the main funders of the US–China collaborative work, to facilitate the investigation into the origins of SARS-CoV-2 (4) has fostered distrust regarding US biodefense research activities.

Much of the work on SARS-like CoVs performed in Wuhan was part of an active and highly collaborative US–China scientific research program funded by the US Government (NIH, Defense Threat Reduction Agency [DTRA], and US Agency for International Development [USAID]), coordinated by researchers at EcoHealth Alliance (EHA), but involving researchers at several other US institutions. For this reason, it is important that US institutions be transparent about any knowledge of the detailed activities that were underway in Wuhan and in the United States. The evidence may also suggest that research institutions in other countries were involved, and those too should be asked to submit relevant information (e.g., with respect to unpublished sequences).

Participating US institutions include the EHA, the University of North Carolina (UNC), the University of California at Davis (UCD), the NIH, and the USAID. Under a series of NIH grants and USAID contracts, EHA coordinated the collection of SARS-like bat CoVs from the field in southwest China and southeast Asia, the sequencing of these viruses, the archiving of these sequences (involving UCD), and the analysis and manipulation of these viruses (notably at UNC). A broad spectrum of coronavirus research work was

done not only in Wuhan (including groups at Wuhan University and the Wuhan CDC, as well as WIV) but also in the United States. The exact details of the fieldwork and laboratory work of the EHA-WIV-UNC partnership, and the engagement of other institutions in the United States and China, has not been disclosed for independent analysis. The precise nature of the experiments that were conducted, including the full array of viruses collected from the field and the subsequent sequencing and manipulation of those viruses, remains unknown.

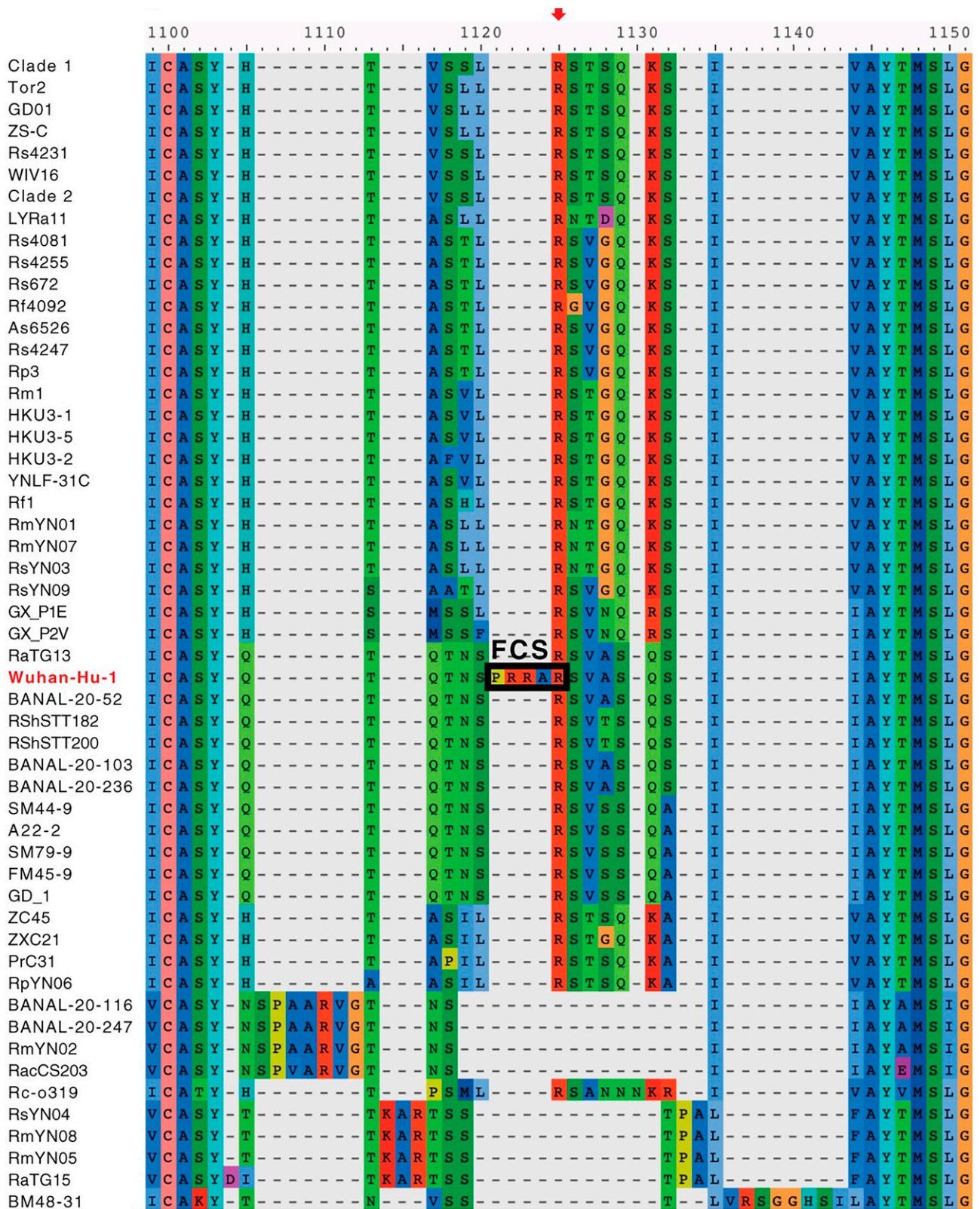
EHA, UNC, NIH, USAID, and other research partners have failed to disclose their activities to the US scientific community and the US public, instead declaring that they were not involved in any experiments that could have resulted in the emergence of SARS-CoV-2. The NIH has specifically stated (6) that there is a significant evolutionary distance between the published viral sequences and that of SARS-CoV-2 and that the pandemic virus could not have resulted from the work sponsored by NIH. Of course, this statement is only as good as the limited data on which it is based, and verification of this claim is dependent on gaining access to any other unpublished viral sequences that are deposited in relevant US and Chinese databases (7,8). On May 11, 2022, Acting NIH Director Lawrence Tabak testified before Congress that several such sequences in a US database were removed from public view, and that this was done at the request of both Chinese and US investigators.

Blanket denials from the NIH are no longer good enough. Although the NIH and USAID have strenuously resisted full disclosure of the details of the EHA-WIV-UNC work program, several documents leaked to the public or released through the Freedom of Information Act (FOIA) have raised concerns. These research proposals make clear that the EHA-WIV-UNC collaboration was involved in the collection of a large number of so-far undocumented SARS-like viruses and was engaged in their manipulation within biological safety level (BSL)-2 and BSL-3 laboratory facilities, raising concerns that an airborne virus might have infected a laboratory worker (9). A variety of scenarios have been discussed by others, including an infection that involved a natural virus collected from the field or perhaps an engineered virus manipulated in one of the laboratories (3).

## Overlooked Details

Special concerns surround the presence of an unusual furin cleavage site (FCS) in SARS-CoV-2 (10) that augments the pathogenicity and transmissibility of the virus relative to related viruses like SARS-CoV-1 (11, 12). SARS-CoV-2 is, to date, the only identified member of the subgenus *sarbecovirus* that contains an FCS, although these are present in other coronaviruses (13, 14). A portion of the sequence of the spike protein of some of these viruses is illustrated in the alignment shown in Fig. 1, illustrating the unusual nature of the FCS and its apparent insertion in SARS-CoV-2 (15). From the first weeks after the genome sequence of SARS-CoV-2 became available, researchers have commented on the unexpected presence of the FCS within SARS-CoV-2—the implication being that SARS-CoV-2 might be a product of laboratory manipulation. In a review

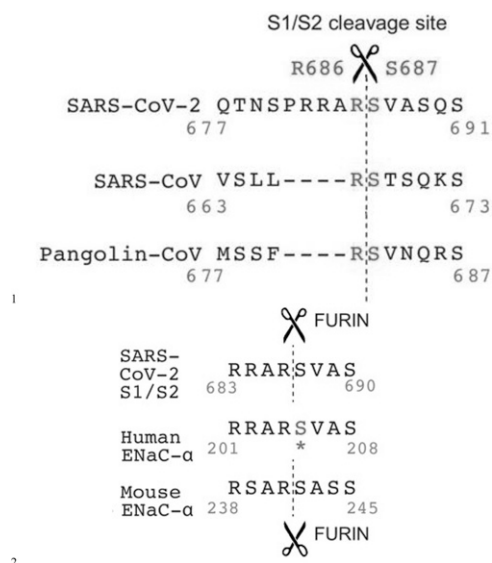




**Fig. 1.** This alignment of the amino acid sequences of coronavirus spike proteins, in the region of the S1/S2 junction, illustrates the sequence of SARS-CoV-2 (Wuhan-Hu-1) and some of its closest relatives. The furin cleavage site (FCS) is indicated (PRRAR'SVAS), and furin cuts the spike protein between R and S, as indicated by the red arrowhead. Adapted from Chan & Zhan (15).

piece arguing against this possibility, it was asserted that the amino acid sequence of the FCS in SARS-CoV-2 is an unusual, nonstandard sequence for an FCS and that nobody in a laboratory would design such a novel FCS (13).

In fact, the assertion that the FCS in SARS-CoV-2 has an unusual, nonstandard amino acid sequence is false. The amino acid sequence of the FCS in SARS-CoV-2 also exists in the human ENaC  $\alpha$  subunit (16), where it is known to be



**Fig. 2. Amino acid alignment of the furin cleavage sites of SARS-CoV-2 spike protein with (Top) the spike proteins of other viruses that lack the furin cleavage site and (Bottom) the furin cleavage sites present in the  $\alpha$  subunits of human and mouse ENaC. Adapted from Anand et al. (16).**

functional and has been extensively studied (17, 18). The FCS of human ENaC  $\alpha$  has the amino acid sequence **RRAR'SVAS** (Fig. 2), an eight-amino-acid sequence that is perfectly identical with the FCS of SARS-CoV-2 (16). ENaC is an epithelial sodium channel, expressed on the apical surface of epithelial cells in the kidney, colon, and airways (19, 20), that plays a critical role in controlling fluid exchange. The ENaC  $\alpha$  subunit has a functional FCS (17, 18) that is essential for ion channel function (19) and has been characterized in a variety of species. The FCS sequence of human ENaC  $\alpha$  (20) is identical in chimpanzee, bonobo, orangutan, and gorilla (*SI Appendix, Fig. 1*), but diverges in all other species, even primates, except one. (The one non-human non-great ape species with the same sequence is *Pipistrellus kuhlii*, a bat species found in Europe and Western Asia; other bat species, including *Rhinolophus ferrumequinem*, have a different FCS sequence in ENaC  $\alpha$  [RKAR'SAAS]).

One consequence of this “molecular mimicry” between the FCS of SARS CoV-2 spike and the FCS of human ENaC is competition for host furin in the lumen of the Golgi apparatus, where the SARS-CoV-2 spike is processed. This results in a decrease in human ENaC expression (21). A decrease in human ENaC expression compromises airway function and has been implicated as a contributing factor in the pathogenesis of COVID-19 (22). Another consequence of this astonishing molecular mimicry is evidenced by apparent cross-reactivity with human ENaC of antibodies from COVID-19 patients, with the highest levels of cross-reacting antibodies directed against this epitope being associated with most severe disease (23).

We do not know whether the insertion of the FCS was the result of natural evolution (2, 13)—perhaps via a recombination event in an intermediate mammal or a human (13, 24)—or was the result of a deliberate introduction of the FCS into a SARS-like virus as part of a laboratory experiment. We do know that the insertion of such FCS sequences into SARS-like viruses was a specific goal of

work proposed by the EHA-WIV-UNC partnership within a 2018 grant proposal (“DEFUSE”) that was submitted to the US Defense Advanced Research Projects Agency (DARPA) (25). The 2018 proposal to DARPA was not funded, but we do not know whether some of the proposed work was subsequently carried out in 2018 or 2019, perhaps using another source of funding.

We also know that that this research team would be familiar with several previous experiments involving the successful insertion of an FCS sequence into SARS-CoV-1 (26) and other coronaviruses, and they had a lot of experience in construction of chimeric SARS-like viruses (27–29). In addition, the research team would also have some familiarity with the FCS sequence and the FCS-dependent activation mechanism of human ENaC  $\alpha$  (19), which was extensively characterized at UNC (17, 18). For a research team assessing the pandemic potential of SARS-related coronaviruses, the FCS of human ENaC—an FCS known to be efficiently cleaved by host furin present in the target location (epithelial cells) of an important target organ (lung), of the target organism (human)—might be a rational, if not obvious, choice of FCS to introduce into a virus to alter its infectivity, in line with other work performed previously.

Of course, the molecular mimicry of ENaC within the SARS-CoV-2 spike protein might be a mere coincidence, although one with a very low probability. The exact FCS sequence present in SARS-CoV-2 has recently been introduced into the spike protein of SARS-CoV-1 in the laboratory, in an elegant series of experiments (12, 30), with predictable consequences in terms of enhanced viral transmissibility and pathogenicity. Obviously, the creation of such SARS-1/2 “chimeras” is an area of some concern for those responsible for present and future regulation of this area of biology. [Note that these experiments in ref. 30 were done in the context of a safe “pseudotyped” virus and thus posed no danger of producing or releasing a novel pathogen.] These simple experiments show that the introduction of the 12 nucleotides that constitute the FCS insertion in SARS-CoV-2 would not be difficult to achieve in a lab. It would therefore seem reasonable to ask that electronic communications and other relevant data from US groups should be made available for scrutiny.

## Seeking Transparency

To date, the federal government, including the NIH, has not done enough to promote public trust and transparency in the science surrounding SARS-CoV-2. A steady trickle of disquieting information has cast a darkening cloud over the agency. The NIH could say more about the possible role of its grantees in the emergence of SARS-CoV-2, yet the agency has failed to reveal to the public the possibility that SARS-CoV-2 emerged from a research-associated event, even though several researchers raised that concern on February 1, 2020, in a phone conversation that was documented by email (5). Those emails were released to the public only through FOIA, and they suggest that the NIH leadership took an early and active role in promoting the “zoonotic hypothesis” and the rejection of the laboratory-associated hypothesis (5). The NIH has

resisted the release of important evidence, such as the grant proposals and project reports of EHA, and has continued to redact materials released under FOIA, including a remarkable 290-page redaction in a recent FOIA release.

Information now held by the research team headed by EHA (7), as well as the communications of that research team with US research funding agencies, including NIH, USAID, DARPA, DTRA, and the Department of Homeland Security, could shed considerable light on the experiments undertaken by the US-funded research team and on the possible relationship, if any, between those experiments and the emergence of SARS-CoV-2. We do not assert that laboratory manipulation was involved in the emergence of SARS-CoV-2, although it is apparent that it could have been. However, we do assert that there has been no independent and transparent scientific scrutiny to date of the full scope of the US-based evidence.

The relevant US-based evidence would include the following information: laboratory notebooks, virus databases, electronic media (emails, other communications), biological samples, viral sequences gathered and held as part of the PREDICT project (7) and other funded programs, and interviews of the EHA-led research team by independent researchers, together with a full record of US agency involvement in funding the research on SARS-like viruses, especially with regard to projects in collaboration with Wuhan-based institutions. We suggest that a bipartisan inquiry should also follow up on the tentative conclusion of the IC (4) that the initial outbreak in Wuhan may have

occurred no later than November 2019 and that therefore the virus was circulating before the cluster of known clinical cases in December. The IC did not reveal the evidence for this statement, nor when parts of the US Government or US-based researchers first became aware of a potential new outbreak. Any available information and knowledge of the earliest days of the outbreak, including viral sequences (8), could shed considerable light on the origins question.

We continue to recognize the tremendous value of US-China cooperation in ongoing efforts to uncover the proximal origins of the pandemic. Much vital information still resides in China, in the laboratories, hospital samples, and early epidemiological information not yet available to the scientific community. Yet a US-based investigation need not wait—there is much to learn from the US institutions that were extensively involved in research that may have contributed to, or documented the emergence of, the SARS-CoV-2 virus. Only an independent and transparent investigation, perhaps as a bipartisan Congressional inquiry, will reveal the information that is needed to enable a thorough scientific process of scrutiny and evaluation.

**ACKNOWLEDGMENTS.** The authors thank Prof. Richard H. Ebright for helpful commentary on the manuscript.

Author affiliations: <sup>a</sup>Department of Anesthesiology, Columbia University, New York, NY 10032; <sup>b</sup>Department of Molecular Pharmacology and Therapeutics, Columbia University, New York, NY 10032; and <sup>c</sup>Center for Sustainable Development, Columbia University, New York, NY 10032

1. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
2. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
3. D. A. Relman, Opinion: To stop the next pandemic, we need to unravel the origins of COVID-19. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 29246–29248 (2020).
4. Office of the Director of National Intelligence, National Intelligence Council, Updated assessment on COVID 19 origins. <https://www.dni.gov/files/ODNI/documents/assessments/Declassified-Assessment-on-COVID-19-Origins.pdf>. Accessed 21 February 2022.
5. M. Hibbett, R. Grim, House republicans release text of redacted Fauci emails on COVID origins. *The Intercept* (12 January 2022). <https://theintercept.com/2022/01/12/covid-origins-fauci-redacted-emails/>. Accessed 21 February 2022.
6. NIAID, SARS-CoV-2 and NIAID-supported bat coronavirus research. An analysis: Evolutionary distance of SARS-CoV-2 and bat coronaviruses studied under the NIH-supported research grant to EcoHealth Alliance (2021). <https://www.niaid.nih.gov/diseases-conditions/coronavirus-bat-research>. Accessed 21 February 2022.
7. S. Suryanarayanan, Ecohealth alliance wanted to block disclosure of covid-19-relevant virus data from China. *US Right to Know* (10 January 2022). <https://usrtrk.org/biohazards-blog/ecohealth-alliance-wanted-to-block-disclosure-of-covid-19-relevant-virus-data-from-china/>. Accessed 21 February 2022.
8. J. D. Bloom, Recovery of deleted deep sequencing data sheds more light on the early Wuhan SARS-CoV-2 epidemic. *Mol. Biol. Evol.* **38**, 5211–5224 (2021).
9. M. Field, A lab assistant involved in COVID-19 research in Taiwan exposed 110 people after becoming infected at work. *Bulletin of the Atomic Scientists* (25 January 2022). <https://thebulletin.org/2022/01/a-lab-assistant-involved-in-covid-19-research-in-taiwan-exposed-110-people-after-becoming-infected-at-work/>. Accessed 21 February 2022.
10. B. Coutard *et al.*, The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* **176**, 104742 (2020).
11. M. Hoffmann, H. Kleine-Weber, S. Pöhlmann, A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol. Cell* **78**, 779–784.e5 (2020).
12. T. P. Peacock *et al.*, The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat. Microbiol.* **6**, 899–909 (2021).
13. E. C. Holmes *et al.*, The origins of SARS-CoV-2: A critical review. *Cell* **184**, 4848–4856 (2021).
14. Y. Wu, S. Zhao, Furin cleavage sites naturally occur in coronaviruses. *Stem Cell Res. (Amst.)* **50**, 102115 (2020).
15. Y. A. Chan, S. H. Chan, The emergence of the spike furin cleavage site in SARS-CoV-2. *Mol. Biol. Evol.* **39**, msab327 (2022).
16. P. Anand, A. Puranik, M. Aravamudan, A. J. Venkatakrishnan, V. Soundararajan, SARS-CoV-2 strategically mimics proteolytic activation of human ENaC. *eLife* **9**, e58603 (2020).
17. A. García-Caballero, Y. Dang, H. He, M. J. Stutts, ENaC proteolytic regulation by channel-activating protease 2. *J. Gen. Physiol.* **132**, 521–535 (2008).
18. P. Kota, M. Gentzsch, Y. L. Dang, R. C. Boucher, M. J. Stutts, The N terminus of  $\alpha$ -ENaC mediates ENaC cleavage and activation by furin. *J. Gen. Physiol.* **150**, 1179–1187 (2018).
19. T. R. Kleyman, D. C. Eaton, Regulating ENaC's gate. *Am. J. Physiol. Cell Physiol.* **318**, C150–C162 (2020).
20. I. Hanukoglu, A. Hanukoglu, Epithelial sodium channel (ENaC) family: Phylogeny, structure-function, tissue distribution, and associated inherited diseases. *Gene* **579**, 95–132 (2016).
21. S. N. Grant, H. A. Lester, Regulation of epithelial sodium channel activity by SARS-CoV-1 and SARS-CoV-2 proteins. *Biophys. J.* **120**, 2805–2813 (2021).
22. M. Gentzsch, B. C. Rossier, A pathophysiological model for COVID-19: Critical importance of transepithelial sodium transport upon airway infection. *Function* **1**, zqaa024 (2020).
23. W. A. Haynes *et al.*, Yale IMPACT Team, High-resolution epitope mapping and characterization of SARS-CoV-2 antibodies in large cohorts of subjects with COVID-19. *Commun. Biol.* **4**, 1317 (2021).
24. W. R. Gallaher, A palindromic RNA sequence as a common breakpoint contributor to copy-choice recombination in SARS-CoV-2. *Arch. Virol.* **165**, 2341–2348 (2020).
25. S. Lerner, M. Hviistendahl, New details emerge about coronavirus research at Chinese lab. *The Intercept* (6 September 2021). <https://theintercept.com/2021/09/06/new-details-emerge-about-coronavirus-research-at-chinese-lab/>. Accessed 22 February 2022.
26. K. E. Folliis, J. York, J. H. Nunberg, Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology* **350**, 358–369 (2006).
27. R. S. Baric, A. C. Sims, Development of mouse hepatitis virus and SARS-CoV infectious cDNA constructs. *Curr. Top. Microbiol. Immunol.* **287**, 229–252 (2005).
28. V. D. Menachery *et al.*, A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **21**, 1508–1513 (2015).
29. B. Yount *et al.*, Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12995–13000 (2003).
30. H. Winstone *et al.*, The polybasic cleavage site in SARS-CoV-2 spike modulates viral sensitivity to type I interferon and IFITM2. *J. Virol.* **95**, e02422-20 (2021).