Dear Editor,

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), emerged in late 2019 and has since caused a pandemic. Although there have been extensive studies worldwide, our understanding of this newly emerged pathogen is far from sufficient. The pathogenesis of the SARS-CoV-2 infection is not fully understood, although a “two-stage” hypothesis was proposed in our previous study.1

As a member of enveloped virus in family Coronaviridae, SARS-CoV-2 makes use of a densely glycosylated spike (S) protein to gain entry into host cells. The S protein is a trimeric class I transmembrane protein composed of two functional subunits. The S1 subunit binds to cellular angiotensin-converting enzyme 2 (ACE2) for host cell recognition, and the S2 subunit functions in viral–host cell membrane fusion. The S protein is the most attractive immunogen for eliciting antibody responses and is therefore the primary focus for neutralizing antibody and vaccine development.

The glycosylation of viral envelope proteins has a wide range of functions, including regulating viral tropism, protein stability and shielding the underlying epitopes from immune surveillance. Thus, a full understanding of the glycosylation of SARS-CoV-2 S protein is critical to reveal the pathogenesis of the virus and to guide the design of therapeutic and prophylactic strategies. A total of 22 N-glycosites were mapped in the in vitro-expressed S protein (Fig. 1a).5 We found that the glycosylation of the S protein extracted from virions is diverse and in sharp difference with the reported glycoforms of purified S protein (Fig. 1b).6 We found that O-glycosylation occurred in clusters on the S protein. The S1 domain was more O-glycosylated with 11 sites, while the remaining 6 sites were detected at the N-terminal of the S2 domain (Fig. 1a). Interestingly, 11 out of 17 identified O-glycosites located near glycosylated Asn, including S60, T124, S151, T236, T604/S605, T618, S659, T1076, T1077, T1097 and T1100 (Fig. 1a, c; Supplementary information, Fig. S3 and S6). The glycopeptide containing T604 and S605 sites was well characterized, however, we were not able to determine the exact glycosylation site due to a lack of diagnostic ions. Therefore, we counted T604/S605 as one O-glycopeptide.

In order to further investigate the dynamics between N- and O-linked glycosylation, we defined the three amino acids on each side of the glycosylated Asn within the consensus motif of NxS/T (x is not proline (P)) as the “position associated to N-sequon” (named N ± 1–3). There were 35 S/T within positions associated to N-seqon; 11 of them were O-glycosylated among which 10 sites were determined. It is intriguing that 7 out of the 10 sites (70%) were located at the N + 2 position, which is in the consensus motif of N-glycosylation (Fig. 1d). All the identified N-glycosites and O-glycosites associated to N-sequon were mapped on the surface of S protein based on the cryo-EM structure of the trimeric SARS-CoV-2 S protein (Protein Data Bank (PDB) ID 6XR8) (Fig. 1e).

To further validate the phenomenon that N- and the O-linked glycosylation occurred together in N-sequon-associated positions, we carried out site-directed mutagenesis. An N-to-Q

Received: 23 March 2021 Accepted: 18 July 2021
Published online: 2 August 2021
mutation was generated on the N-linked glycosite N616 on purified full-length WT S protein. The O-glycosite T618 was analyzed together with the deamidated N616 (Supplementary information, Fig. S4). Mutations of N616 completely abolished the O-glycosylation on T618 (Fig. 1f), indicating that the presence of glycosylated Asn is the prerequisite of O-glycosylation associated to N-sequon.

Based on the observations above, we proposed an “O-Follow-N” rule, whereby O-glycosylation occurs near the glycosylated Asn in N-sequon. This may also apply to other proteins and promote the
identification of O-glycosites (Fig. 1g). It has been reported that GalNAc-transferases (GalNAc-Ts), which mediate the initiation of mucin-type O-glycosylation, contain a lectin-like domain that binds glycans. We reasoned that GalNAc-Ts may recognize glycans on the N-glycosites, thus catalyzing O-glycosylation near the N-glycosites.

In summary, we conducted a site-specific glycosylation analysis, and comprehensively profiled the N- and O-linked glycosylation of the S protein either extracted from virions or in vitro expressed. To our knowledge, this is the first and largest glycosylation dataset of S protein directly extracted from the SARS-CoV-2 virions to date, which broadens our understanding of the glycosylation of the SARS-CoV-2 S protein. In this context, we observed a unique pattern that to our knowledge, this is the first and largest glycosylation dataset of S protein directly extracted from the SARS-CoV-2 virions to date, which broadens our understanding of the glycosylation of the SARS-CoV-2 S protein. In this context, we observed a unique pattern that

O followed by N (O-Follow-N) rule has been proposed due to the limited number of identifications. The observation reported here of O-glycosites in close proximity to N-glycosylation suggests the possible "O-Follow-N" rule. It has long been known that N-glycosylation occurs in the NXS/T (x is not proline (P)) consensus motif, however, it is not clear whether the S/T after glycosylated Asn is prone to be O-glycosylated. If this is the case, it would be interesting to explore whether the glycosylation of S/T depends on the nearby N-glycosylation. The "O-Follow-N" rule discovered in this study would shed light on the potential new mechanisms of O-glycosylation, especially the synergies between N- and O-glycosylation, and would greatly benefit fundamental glyobiology studies.

DATA AVAILABILITY
The mass spectrometry raw files have been deposited in the MassIVE proteomics database under the accession number PXD023346.

REFERENCES
1. Tian, W. et al. Nat. Commun. 11, 5859 (2020).
2. Watanabe, Y., Allen, J. D., Wrapp, D., McElhan, J. S. & Crispin, M. Science 369, 330–333 (2020).
3. Shajahan, A., Supekar, N. T., Gleinich, A. S. & Azadi, P. Glycobiology 30, 981–988 (2020).
4. Sanda, M., Morrison, L. & Goldman, R. Anal. Chem. 93, 2003–2009 (2021).
5. Bagdonaitė, I. et al. Viruses 13, 551 (2021).
6. Xu, T., Wong, C. C., Kashina, A. & Yates, J. R. 3rd Nat. Protoc. 4, 325–332 (2009).
7. Yao, H. et al. Cell 183, 730–748.e13 (2020).
8. Imberty, A., Piller, V., Piller, F. & Breton, C. Protein Eng. 10, 1353–1356 (1997).
9. Riethmüller, S. et al. PLoS Biol. 15, e2000080 (2017).
10. Chandrasekhar, K. D. et al. J. Physiol. 589, 3721–3730 (2011).

ACKNOWLEDGEMENTS
We would like to thank Min Huang, Yue Zhou, Guanbo Wang, Minjie Tan for their assistance in mass spectrometry experiments. We would also like to thank Thermo Fisher Scientific for their support in collaboration. We would like to thank Shuaxin Gao, Yu Hu, Chao Su, Qihui Wang for their assistance in S protein purification. We thank Hui Wang and Beijing Institute of Biological Products Co., Ltd. for providing SARS-CoV-2 virions. We are grateful to Xuefei Yin (PML Inc.), Yan Zhang, Wenfeng Zeng and Mingliang Ye for their insightful discussion. This work was supported by the Fundamental Research Funds for the Central Universities (BMU2017Y003, BMU2018XZ2002), the Research Funds from Health@InnoHK Program launched by the Innovation Technology Commission of the Hong Kong Special Administrative Region, the PKU-Baidu Fund (2019BD0007), the Training Program of the Big Science Strategy Plan (2020YFE0202200), the National Program on Key Research Projects of China (2016YFD0503001), and the Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICP-KJGG-004).

AUTHOR CONTRIBUTIONS
G.F.G. and C.C.L.W. conceived the project. D.L., G.B., K.Y., and H.X. expressed and purified the proteins. C.C.L.W. and W.T. supervised the mass spectrometry proteomics experiments. N.Z., performed the mass spectrometry experiments. C.C.L.W., G.F.G., W.T., Y.C., N.Z., F.G., and D.L. analyzed the data. G.F.G., C.C.L.W., Y.C., and W.T. wrote the manuscript.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41422-021-00545-2.

Correspondence and requests for materials should be addressed to F.G., Y.C., C.C.L.W., or G.F.G.

Reprints and permission information is available at http://www.nature.com/reprints

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021