COVIDep: a web-based platform for real-time reporting of vaccine target recommendations for SARS-CoV-2

To the Editor — The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has brought much of the world to a virtual lockdown. As the virus continues to spread rapidly and the pandemic intensifies, the need for an effective vaccine is becoming increasingly apparent. A critical part of vaccine design is to identify targets, or epitopes, that can induce an effective immune response against SARS-CoV-2. This process is challenged by our limited understanding of this novel coronavirus and of its interplay with the human immune system.

In response to this challenge, we have developed COVIDep (COVIDep.ust.hk), a first-of-its-kind web-based platform that pools genetic data for SARS-CoV-2 and immunological data for the 2003 SARS virus, SARS-CoV, to identify B-cell and T-cell epitopes to serve as vaccine target recommendations for SARS-CoV-2 (Fig. 1a). For T-cell epitopes, it provides estimates of population coverage, globally and for specific regions. COVIDep is flexible and user-friendly, comprising an intuitive graphical interface and interactive visualizations. In addition to producing formatted, exportable lists of the identified B-cell and T-cell epitopes and their basic characteristics, COVIDep includes displays for each of the SARS-CoV-2 proteins, showing the locations of the identified epitopes on the primary structure. Further graphical displays are provided to aid interpretation of the data, including a temporal and geographical breakdown of the analyzed sequences, and a display of the observed genetic variation (amino acid mutation frequencies) for each SARS-CoV-2 protein. The platform is updated daily, based on the latest SARS-CoV-2 sequence data in the GISAID database (www.gisaid.org). Periodic updates are important because SARS-CoV-2 sequences are being made available at an increasing rate through international data-sharing efforts, and the identification of vaccine targets is influenced by newly observed genetic variation.

The vaccine targets recommended by COVIDep exploit the genetic similarities between SARS-CoV-2 and SARS-CoV, along with known immune targets for SARS-CoV that have been determined experimentally (available in the ViPR database; www.viprbrc.org). The system implements a protocol that identifies, from among the SARS epitopes that can induce a human immune response, those that are genetically similar in SARS-CoV-2. This approach, proposed and tested in our preliminary study[1] based on limited early data, identified known SARS-CoV epitopes that had an identical genetic match in SARS-CoV-2. These epitopes presented initial vaccine target recommendations for potentially eliciting a protective, cross-reactive immune response against SARS-CoV-2. Similar results were reported in a subsequent independent study[2], in which a related approach exploiting genetic similarity between SARS-CoV and SARS-CoV-2 was used to identify potential SARS-CoV-2 vaccine targets.

The use of SARS-CoV immunological data to inform vaccine targets for SARS-CoV-2 is being supported by experimental results. There is evidence of cross-neutralization by SARS-CoV-derived antibodies binding to genetically similar regions of SARS-CoV-2's spike protein[3–5]. Conversely, studies have demonstrated that specific SARS-CoV-derived antibodies that bind to the spike's receptor-binding domain, which has significant genetic differences in SARS-CoV-2, have limited cross-reactivity[6]. T cell responses against spike protein epitopes that are genetically similar in SARS-CoV and SARS-CoV-2 have also been reported in COVID-19-infected patients[7,8] and in a preclinical vaccine trial[9] (Fig. 1b). Epitopes recommended by COVIDep have notable overlap with the findings in these and other[10,11] experimental studies (see Figs. 2 and 3 in ref. [12]). The recommendations provided by COVIDep may be used to broadly guide vaccine designs and associated experimental studies, and may help to expedite the discovery of an effective vaccine for COVID-19. ❐

Editor's note: This article has been peer-reviewed.

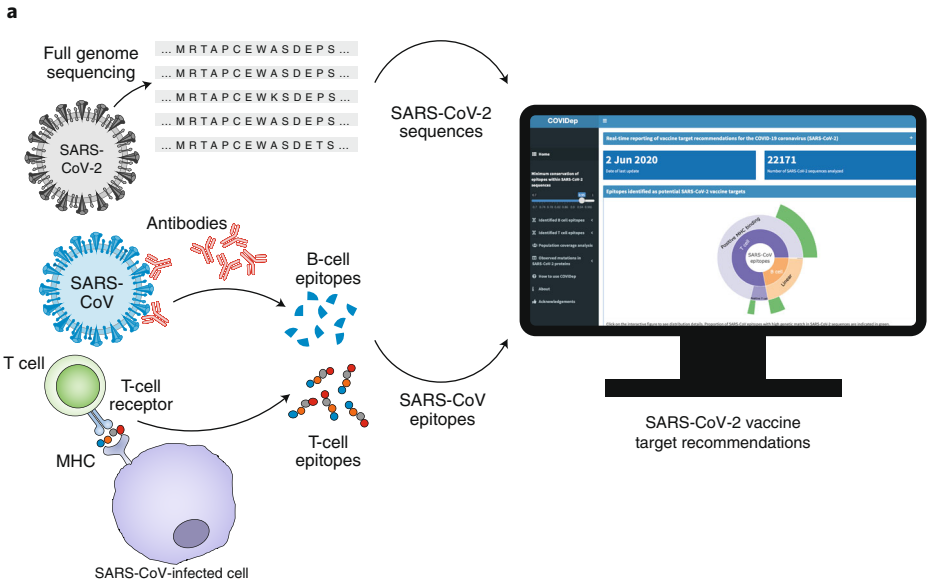
Syed Faraz Ahmed [1], Ahmed A. Quadeer [1]✉ and Matthew R. McKay [1,2]✉

[1]*Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.* [2]*Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.* ✉*e-mail: eeaaquadeer@ust.hk; m.mckay@ust.hk*

Published online: 17 June 2020
https://doi.org/10.1038/s41596-020-0358-9

References

1. Ahmed, S. F., Quadeer, A. A. & McKay, M. R. *Viruses* **12**, 254 (2020).
2. Grifoni, A. et al. *Cell Host Microbe* **27**, 1–10 (2020).
3. Walls, A. C. et al. *Cell* **180**, 1–12 (2020).
4. Wang, C. et al. *Nat. Commun.* **11**, 2251 (2020).
5. Pinto, D. et al. *Nature* https://doi.org/10.1038/s41586-020-2349-y (2020).
6. Wrapp, D. et al. *Science* **367**, 1260–1263 (2020).
7. Chour, W. et al. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.04.20085779 (2020).
8. Shomuradova, A. S. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2020.05.20.20107813 (2020).
9. Smith, T. R. F. et al. *Nat. Commun.* **11**, 2601 (2020).
10. Poh, C. M. et al. *Nat. Commun.* **11**, 2806 (2020).
11. Yin, D. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2020.05.14.093054 (2020).
12. Ahmed, S. F., Quadeer, A. A. & McKay, M. R. Preprint at *bioRxiv* https://doi.org/10.1101/2020.05.23.111385 (2020).



b

Details of the identified T cell epitopes in the S protein

Show 25 entries

Download csv

Search: HLA-A*02:01

IEDB	Epitope	Length	Start	End	MHC allele class	MHC allele names	Conservation
36724	LITGRLQSL	9	996	1004	I	HLA-A2/HLA-A*02:01	0.9998
54507	RLDKVEAEV	9	983	991	I	HLA-A*02:01/HLA-A*02:02/HLA-A*02:06/HLA-A*02:03/HLA-A*68:02	0.9998
54725	RLQSLQTYV	9	1000	1008	I	HLA-A*02:01/HLA-A*02:02/HLA-A*02:03/HLA-A*02:06/HLA-A*68:02	0.9998
37544	LLLQYGSFC	9	752	760	I	HLA-A*02:01	0.9997
37724	LLQYGSFCT	9	753	761	I	HLA-A*02:01	0.9997
69657	VLNDILSRL	9	976	984	I	HLA-A*02:01	0.9997
2801	ALNTLVKQL	9	958	966	I	HLA-A*02:01	0.9994
71663	VVFLHVTYV	9	1060	1068	I	HLA-A*02:01/HLA-A*02:02/HLA-A*02:03/HLA-A*02:06/HLA-A*68:02	0.9994
44814	NLNESLIDL	9	1192	1200	I	HLA-A*02:01	0.9993
26710	IITTDNTFV	9	1114	1122	I	HLA-A*02:01	0.9992
54680	RLNEVAKNL	9	1185	1193	I	HLA-A*02:01	0.9992
16156	FIAGLIAIV	9	1220	1228	I	HLA-A*02:01/HLA-A*02:02/HLA-A*02:03/HLA-A*02:06/HLA-A*68:02/HLA-A2	0.9991
20907	GLIAIVMVTI	10	1223	1232	I	HLA-A*02:02/HLA-A*02:03/HLA-A*02:01/HLA-A*02:06/HLA-A*68:02	0.9984
37289	LLFNKVTLA	9	821	829	I	HLA-A*02:01/HLA-A*02:02/HLA-A*02:03/HLA-A*02:06/HLA-A*68:02	0.9976

Showing 1 to 14 of 14 entries (filtered from 75 total entries)

Previous 1 Next

Fig. 1 | COVIDep provides an up-to-date set of B-cell and T-cell epitopes that can serve as potential vaccine targets for SARS-CoV-2. a, The identified epitopes are experimentally derived from SARS-CoV and have a close genetic match with the available SARS-CoV-2 sequences (see Supplementary Figure 1 for a detailed protocol description). b, An example of the T-cell epitopes reported by COVIDep (as of 20 May 2020) for the spike protein of SARS-CoV-2. Here, the Search box (in the top right) was used to select only the HLA-A*02:01-restricted epitopes. (An explanation of all interactive COVIDep visualizations is incorporated in the 'How to use COVIDep' page of the platform.) Of the 14 epitopes listed in the display, 9 (IEDB IDs 36724, 54507, 54725, 69657, 71663, 2801, 54680, 16156 and 37289) overlap with epitopes against which cytotoxic $CD8^+$ T-cell responses have been observed in peripheral blood mononuclear cells isolated from COVID-19 patients[7,8]. T-cell responses were also recorded against protein regions overlapping with the epitope with IEDB ID 71663 in a preclinical trial of a DNA vaccine candidate[9].

Acknowledgements

COVIDep is made possible by the open sharing of genome sequence data of SARS-CoV-2 sequences by research groups from around the world through the GISAID platform, and the open sharing of immunological data of experimentally determined SARS-CoV epitopes through the ViPR database. We gratefully acknowledge the contributions of all the researchers, scientists and technical staff involved (a detailed acknowledgment is available at the Acknowledgments page of the COVIDep platform). We thank N. Law and K. Pang for their technical support, and R. Louie, D. Morales-Jimenez, S. Sohail, N. Kundu, A. Shah and U. Abdullah for comments and suggestions that helped to improve the web application and its interface. M.R.M. and A.A.Q. were supported by the General Research Fund of the Hong Kong Research Grants Council (RGC) (grant 16204519). S.F.A. was supported by the Hong Kong Ph.D. Fellowship Scheme (HKPFS).

Data availability

The SARS-CoV-2 full genome sequence data used in COVIDep is periodically downloaded from the Global Initiative on Sharing Avian Influenza Database (GISAID; www.gisaid.org). The SARS-CoV epitope sequence data was downloaded from the Virus Pathogen Database and Analysis Resource (ViPR; www.viprbrc.org). The population coverage statistics of HLA alleles were obtained from the Immune Epitope Database and Analysis Resource (IEDB; www.iedb.org).

Code availability

The source code for the developed platform is available at the COVIDep GitHub repository (github.com/COVIDep).

Competing interests

The authors declare no competing interests.