

COVIEdb: A database for potential immune epitopes of coronaviruses

Jingcheng Wu^{1,2}, Wenfan Chen¹, Jingjing Zhou¹, Wenyi Zhao^{1,2}, Shuqing Chen¹, Zhan Zhou^{1,3,*}

¹ Institute of Drug Metabolism and Pharmaceutical Analysis and Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

² College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China

³ Innovation Institute for Artificial Intelligence in Medicine, Zhejiang University, Hangzhou 310018, China

*Corresponding authors.

E-mail: zhanzhou@zju.edu.cn (Zhou Z).

Abstract

2019 novel coronavirus (2019-nCoV) has caused large-scale pandemic COVID-19 all over the world. It's essential to find out which parts of the 2019-nCoV sequence are recognized by human immune system for vaccine development. And for the prevention of the potential outbreak of similar coronaviruses in the future, vaccines against immunogenic epitopes shared by different human coronaviruses are essential. Here we predict all the potential B/T-cell epitopes for SARS-CoV, MERS-CoV, 2019-nCoV and RaTG13-CoV based on the protein sequences. We found YFKYWDQTY in ORF1ab protein, VYDPLQPEL and TVYDPLQPEL in spike (S) protein might be pan-coronavirus targets for vaccine development. All the predicted results are stored in a database COVIEDb (<http://biopharm.zju.edu.cn/coviedb/>).

Keywords: Coronavirus, Epitopes, Vaccine, Database, COVID-19

Introduction

Two coronaviruses—severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) have caused two large-scale pandemics in the past two decades [1,2]. Now, the third coronavirus caused pandemic (COVID-19) is ongoing [3,4]. The 2019 novel coronavirus (2019-nCoV) which was first identified in Wuhan, China in December 2019, from patients with pneumonia is the very coronavirus [5]. Analysis of the viral genome has revealed that 2019-nCoV is phylogenetically close to SARS-CoV [6], as was named SARS-CoV-2. As of April 27, 2020, 2,980,053 people have been confirmed COVID-19, including 206,569 deaths (~6.9% fatality rate) all over the world.

Because of the less cost-effective than treatment, and reduce morbidity and mortality without long-lasting effects, vaccines are the most effective strategy for preventing infectious diseases [7]. However, there is still no approved vaccines for human coronaviruses (hCoV). There are several types of vaccines are under pre-clinical testing or clinical trials including inactivated vaccine, recombinant subunit vaccine, recombinant vector vaccine, and nucleic acid vaccine. In general, modern vaccines,

such as recombinant subunit, peptide, and nucleic acid vaccines, are advantageous over other types of vaccines because of higher safety and less side effect, by inducing the immune system without introducing whole infectious viruses [8]. Nucleic acid vaccines such as DNA vaccines and mRNA vaccines represent an innovative approach by direct injection of plasmids or mRNAs encoding the antigens, accompanied with a wide range of immune responses [9,10]. These advantages are applied with prophylactic vaccines and therapeutic vaccines to treat infectious diseases and cancers. For the development of modern vaccines, it is of critical importance to identify potential immune epitopes of 2019-nCoV, as well as other infectious pathogens.

Considering the seriousness of the recent outbreaks of zoonotic coronaviruses, therapeutic agents and vaccines for pan-coronaviruses should be developed to cope with the hCoV outbreaks in the present and in the future. Here we predict all the potential B/T cell epitopes for SARS-CoV, MERS-CoV and 2019-nCoV to provide potential targets for pan-coronaviruses vaccine development. The prediction are based on their proteins sequences. RaTG13-CoV is added because of its high homology with 2019-nCoV (96% whole genome identity [11]). All the predicted results are stored in a database named COVIEdb (<http://biopharm.zju.edu.cn/coviedb/>).

Database content and usage

Data Source

The protein sequences of SARS-CoV, MERS-CoV, 2019-nCoV, and RaTG13-CoV are downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/>). The genome sequences of SARS-CoV, MERS-CoV, 2019-nCoV and RaTG13 encode 13, 10, 11 and 10 proteins, respectively. Among these protein sequences, ORF1ab is not divided. And the N protein of SARS-CoV is divided into ORF9a and ORF9b.

The human leukocyte antigen (HLA) alleles used for T-cell epitopes prediction are derived from Zhou *et.al.* which analyzed the HLA distribution of 20,635 individuals of Han Chinese ancestry [12]. The top 20 HLA I alleles of A, B, C subtypes and HLA II alleles with frequency more than 5% are the final HLA datasets (Table S1).

B-cell epitopes prediction

The B-cell epitopes were predicted by the seven tools embedded in the Immune Epitope Database and Analysis Resource (IEDB) [13]. More specifically, BepiPred-1.0 [14], BepiPred-2.0 [15], Chou and Fasman beta turn prediction [16], Emini surface accessibility scale [17], Karplus and Schulz flexibility scale [18], Kolaskar and Tongaonkar antigenicity scale [19] and Parker hydrophilicity prediction [20] are used for predicting amino acid sites belonging to B-cell epitopes. The parameters are all set as default. The sites confirmed by at least five tools would be labeled in red. The thresholds of each tool are in Table S2a. In this study, only amino acids that be confirmed by at least five tools are considered as B-cell epitopes.

T-cell epitopes prediction

The T-cell epitopes prediction were divided into two parts. One of them are presented by HLA I allele and would induce the activation of CD8+ T cells. This type of T-cell epitopes were predicted by NetMHCpan 4.0 [21], MHCflurry [22] and DeepHLApan [23]. Another type of T-cell epitopes presented by HLA II alleles were predicted by MixMHC2pred [24] and NetMHCIIpan [25], which would induce the activation of CD4+ T cells. The thresholds to defined potential T-cell epitopes of each tool are listed in Table S2b.

In the prediction of T-cell epitopes presented by HLA I alleles, all peptides with length range from 8 to 11 were selected and combined with previous HLA I alleles to create HLA-peptide pairs. It's similar to predict that presented by HLA II alleles, with the difference that peptide length range from 15 to 28. Only HLA-peptide pairs satisfied with all thresholds of used tools would be potential T-cell epitopes in this study.

Web interface

The web interface of COVIEdb is made up of five parts, “Home”, “B-epitope”, “T-epitope”, “Peptide”, and “Help”.

A statistical table about the predicted B-cell epitopes and T-cell epitopes are embedded in the “Home” page. The table shows the coronavirus, gene, protein length, the number of amino acids that predicted to be potential B-cell epitopes and T-cell epitopes (Table 1). The number of the predicted epitopes are different but similar among four coronaviruses. The N protein of SARS-CoV is represented as ORF9a and ORF9b

as previous state. We can find that many peptides have the potential to be B-cell epitopes or T-cell epitopes, but the prefer targets should be shared by different coronaviruses and we will discuss in the next section.

The predicted results of B-cell epitopes could be searched in “B-epitope” page. With the virus and gene selected, the corresponding predicted B-cell epitopes would appear (Figure 1a is the example result when the selected coronavirus and protein are 2019-nCoV and S, respectively). The first to fourth columns are the basic information of the amino acids in coronavirus. The following seven columns are the results that different tools predicted if it’s the amino acid in B-cell epitopes. Once click the blue number, the corresponding score and threshold would show (Figure 1b). The last column is the number of tools that the amino acid is predicted as potential B-cell epitopes. On click the number, the corresponding tools would show (Figure 1c).

The predicted results of T-cell epitopes could be searched in “T-epitope” page. Similar with that in “B-epitope” page, coronavirus and protein are necessary. Besides, the type of T-cell epitopes should also be selected. Because of the numerous pairs of HLA-peptide, we don’t show predicted score of all pairs. Only the peptide-HLA pairs which satisfied all thresholds would be showed in this page. Figure 2a and Figure 2b show the searching results when selecting “2019-nCoV, S and HLA I allele”, “2019-nCoV, S and HLA II allele”, respectively.

The searchable data in the “Peptide” page is the combined result of previous predicted B-cell epitopes and T-cell epitopes. In this page, the only selectable parameter is the protein. On click the search button with protein S selected, a table will show (Figure 3a). The first column is the protein name. The second column is the peptide that either be predicted T-cell epitopes presented by HLA I alleles (with green background) or HLA II alleles (with yellow background). Amino acids that have been predicted as B cell epitopes by at least five tools are labeled in red. The third is an adjustment score to evaluate the possibility of this peptide to be B cell epitopes. The formulation is as follows:

$$B_score = \frac{\sum n_a}{L}$$

Where L is the length of the peptide, a is the amino acid that belongs to the peptide, n_a is the number of tools convinced that amino acid is part of B-cell epitopes. The fourth/fifth columns are the frequency of HLA I/II alleles that could present the peptide. On click the number would show the detailed HLA alleles (Figure 3b). The last column is the number of coronavirus that encodes the peptide. The detailed coronavirus and the peptide location in this coronavirus would show on click the number (Figure 3c). The table displays in descent according to the frequency of HLA I alleles by default. However, other columns could also be sortable.

Shared B/T-cell epitopes

Though the evolution rate of human coronavirus is fast, we try to find out B/T cell epitopes conserved and shared in different coronaviruses for the pan-coronavirus vaccine development. Based on the predicted B-cell epitopes and T-cell epitopes, we found 77 peptides that exist in all coronaviruses have the potential to induce T-cell activation and 10 of them with B_score larger than 4 (Table S3, Table 2). In particular, the peptide YFKYWDQTY from ORF1ab could be presented by 7.33% people, which might be a good candidate for vaccine design.

All the T-cell epitopes shared in four coronaviruses are located in ORF1ab. However, the S protein of the coronavirus is the most important protein where the receptor binding domain (RBD) located. So we further investigated the shared epitopes that located in S protein. There are 265 potential epitopes in S protein shared by three coronaviruses and 35 of them with B_score larger than 5 (Table S4). The peptides VYDPLQPEL and TVYDPLQPEL even have B_score larger than 6, and they can be presented by 8.26% and 2.44% people, respectively, which indicates their feasibility to be pan-coronavirus vaccine targets.

The shared B/T-cell epitopes could be found in the “Peptide” page in COVIEDb and we believe that these results could benefit the vaccine development currently on 2019-nCoV and also the potential coronavirus outbreak in the future.

Discussion

Research groups all over the world are trying to find the way to control 2019-nCoV. Several studies focus on the B/T-cell epitopes for accelerating vaccine development. The group who maintain the database IEDB identified potential targets for immune responses to the 2019-nCoV by sequence homology with closely related SARS-CoV and by a priori epitope prediction using bioinformatics approaches [26]. A recent accepted work executed a comprehensive *in silico* analysis of viral peptide-HLA I allele binding affinity across 145 HLA -A, -B, and -C genotypes for all 2019-nCoV peptides to discover the HLA susceptibility map for 2019-nCoV [27].

However, the vaccine development can't react immediately to control a novel coronavirus epidemic. The better choice is to prepare candidate vaccines before the outbreak coming. For this purpose, we predicted all the potential B/T-cell epitopes of SARS-CoV, MERS-CoV, 2019-nCoV and RaTG13-CoV and want to find out the shared B/T-cell epitopes among them, which may be potential pan-coronavirus vaccine targets. Based on the predicted results by different tools, we find several peptides that could be vaccine candidates for hCoVs such as YFKYWDQTY in ORF1ab, VYDPLQPEL and TVYDPLQPEL in S protein. We believe that these results and the developed database could benefit not only the vaccine (especially the polyvalent vaccine which could protect from various coronavirus) development but also provide the targets for drug design such as neutralizing antibody on 2019-nCoV and the possible coronavirus outbreak in the future.

Acknowledgment

This work was supported by the Key R&D Program of Zhejiang Province (Grant No. 2020C03010), the National Natural Science Foundation of China (Grant No. 31971371), the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY19H300003), and the Fundamental Research Funds for the Central Universities of China.

References

- [1] Drosten C, Günther S, Preiser W, Van der Werf S, Brodt HR, Becker S, et al.

Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N

- Engl J Med 2003;348:1967–76. <https://doi.org/10.1056/NEJMoa030747>.
- [2] Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 2012;367:1814–20. <https://doi.org/10.1056/NEJMoa1211721>.
- [3] Zhang N, Wang L, Deng X, Liang R, Su M, He C, et al. Recent advances in the detection of respiratory virus infection in humans. *J Med Virol* 2020;92:408–17. <https://doi.org/10.1002/jmv.25674>.
- [4] Liu SL, Saif L. Emerging Viruses without Borders: The Wuhan Coronavirus. *Viruses* 2020;12:E130. <https://doi.org/10.3390/v12020130>.
- [5] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727–33. <https://doi.org/10.1056/NEJMoa2001017>.
- [6] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395:565–74. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8).
- [7] Zhang C, Maruggi G, Shan H, Li J. Advances in mRNA vaccines for infectious diseases. *Front Immunol* 2019;10:1–13. <https://doi.org/10.3389/fimmu.2019.00594>.
- [8] Graham RL, Donaldson EF, Baric RS. A decade after SARS: Strategies for controlling emerging coronaviruses. *Nat Rev Microbiol* 2013;11:836–48. <https://doi.org/10.1038/nrmicro3143>.
- [9] Yang ZY, Kong WP, Huang Y, Roberts A, Murphy BR, Subbarao K, et al. A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. *Nature* 2004;428:561–4. <https://doi.org/10.1038/nature02463>.
- [10] Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines—a new era in vaccinology. *Nat Rev Drug Discov* 2018;17:261–79. <https://doi.org/10.1038/nrd.2017.243>.
- [11] Zhou P, Yang X Lou, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579:270–3. <https://doi.org/10.1038/s41586-020-2012-7>.
- [12] Zhou F, Cao H, Zuo X, Zhang T, Zhang X, Liu X, et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. *Nat Genet*

- 2016;48:740–6. <https://doi.org/10.1038/ng.3576>.
- [13] Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 2015;43:D405–12. <https://doi.org/10.1093/nar/gku938>.
- [14] Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res* 2006;2:2. <https://doi.org/10.1186/1745-7580-2-2>.
- [15] Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017;45:W24–9. <https://doi.org/10.1093/nar/gkx346>.
- [16] Enzymology AIN. Prediction of the secondary structure of proteins from their amino acid sequence. vol. 47. 1978.
- [17] Emini EA, Hughes J V, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 1985;55:836–9. <https://doi.org/10.1128/jvi.55.3.836-839.1985>.
- [18] Karplus PA, Schulz GE. Prediction of chain flexibility in proteins - A tool for the selection of peptide antigens. *Naturwissenschaften* 1985;72:212–3. <https://doi.org/10.1007/BF01195768>.
- [19] Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990;276:172–4. [https://doi.org/10.1016/0014-5793\(90\)80535-Q](https://doi.org/10.1016/0014-5793(90)80535-Q).
- [20] Parker JMR, Guo D, Hodges RS. New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-ray-Derived Accessible Sites. *Biochemistry* 1986;25:5425–32. <https://doi.org/10.1021/bi00367a013>.
- [21] Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* 2017;199:3360–8. <https://doi.org/10.4049/jimmunol.1700893>.
- [22] Donnell TJO, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J, et al. MHCflurry : Open-Source Class I MHC Binding Tool. *Cell Syst* 2018;7:129-132.e4.

- <https://doi.org/10.1016/j.cels.2018.05.014>.
- [23] Wu J, Wang W, Zhang J, Zhou B, Zhao W, Su Z, et al. DeepHLApan: A deep learning approach for neoantigen prediction considering both HLA-peptide binding and immunogenicity. *Front Immunol* 2019;10:2559.
<https://doi.org/10.3389/fimmu.2019.02559>.
- [24] Chen B, Khodadoust MS, Olsson N, Wagar LE, Fast E, Liu CL, et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat Biotechnol* 2019;37:1332–43. <https://doi.org/10.1038/s41587-019-0280-2>.
- [25] Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics* 2013;65:711–24.
<https://doi.org/10.1007/s00251-013-0720-y>.
- [26] Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* 2020:1–10.
<https://doi.org/10.1016/j.chom.2020.03.002>.
- [27] Lee S, Hahn C, Rhee M, Oh JE, Song J, Chen Y, et al. Human leukocyte antigen susceptibility map for SARS-CoV-2. *J Virol* 2020;accept.
<https://doi.org/10.1017/CBO9781107415324.004>.

Figure legends

Figure 1. The searching result of “B-epitope” page when selecting 2019-nCoV and S. **a.** the whole interface. **b.** the predicted score and corresponding threshold on click of blue number in the column of seven tools. **c.** the detailed tools that convince the amino acids is part of B-cell epitopes.

Figure 2. a. The searching result of “T-epitope” page when selecting 2019-nCoV, S and HLA-I allele. **b.** The searching result of “T-epitope” page when selecting 2019-nCoV, S and HLA-II allele.

Figure 3. a. The searching result of “Peptide” page when selecting S. **b.** the detailed HLA alleles that could present peptides. **c.** The detailed coronavirus and the peptide location in this coronavirus when click the number in the last column.

Tables

Table 1. The potential immune epitopes from E, M, N, and S proteins for each coronavirus in the database.

Coronavirus	Gene	Length	No. of amino acids for B-cell epitopes	No. of T-cell epitopes presented by HLA I alleles	No. of T-cell epitopes presented by HLA II alleles
2019-nCoV	E	75	21	40	68
MERS-CoV	E	82	9	29	76
RaTG13-CoV	E	75	19	40	68
SARS-CoV	E	76	18	37	56
2019-nCoV	M	222	70	125	333
MERS-CoV	M	219	43	134	246
RaTG13-CoV	M	221	41	125	329
SARS-CoV	M	221	45	133	346
2019-nCoV	N	419	155	90	244
MERS-CoV	N	413	184	105	204
RaTG13-CoV	N	419	191	86	238
SARS-CoV	ORF9a	422	188	98	231
SARS-CoV	ORF9b	98	30	25	47
2019-nCoV	S	1273	359	431	1471
MERS-CoV	S	1353	357	482	1804
RaTG13-CoV	S	1278	297	423	1621
SARS-CoV	S	1255	279	441	1564

Table 2. The potential T-cell epitopes with B_score larger than 4.

Gene	Peptide	HLA able to present peptide	Peptide location in different coronaviruses	B_score	HLA I frequency	HLA II frequency
ORF1ab	KPGGTSSGDATTAYA	DQA1_05_05__DQB1_03_01	SARS-CoV:5045_5059 2019-nCoV:5068_5082 MERS-CoV:5054_5068 RaTG13-CoV:5067_5081	4.98	0	0.42
ORF1ab	TSSGDATTAY	HLA-B15:01	2019-nCoV:5072_5081 RaTG13-CoV:5071_5080 SARS-CoV:5049_5058 MERS-CoV:5058_5067	4.98	1.68	0
ORF1ab	SSGDATTAY	HLA-B15:01 HLA-B35:01 HLA-B15:02	2019-nCoV:5073_5081 RaTG13-CoV:5072_5080 SARS-CoV:5050_5058 MERS-CoV:5059_5067	4.97	3.4	0
ORF1ab	KPGGTSSGDATTAYAN	DQA1_05_05__DQB1_03_01	SARS-CoV:5045_5060 2019-nCoV:5068_5083 MERS-CoV:5054_5069 RaTG13-CoV:5067_5082	4.73	0	0.42
ORF1ab	GGTSSGDATTAYAN	DQA1_05_05__DQB1_03_01	SARS-CoV:5047_5060 2019-nCoV:5070_5083 MERS-CoV:5056_5069 RaTG13-CoV:5069_5082	4.7	0	0.42
ORF1ab	GGTSSGDATTAYANS	DQA1_05_05__DQB1_03_01	SARS-CoV:5047_5061 2019-nCoV:5070_5084 MERS-CoV:5056_5070 RaTG13-CoV:5069_5083	4.47	0	0.42
ORF1ab	GTSSGDATTAYANS	DQA1_05_05__DQB1_03_01	SARS-CoV:5048_5061 2019-nCoV:5071_5084 MERS-CoV:5057_5070 RaTG13-CoV:5070_5083	4.36	0	0.42
ORF1ab	GGTSSGDATTAYANSV	DQA1_05_05__DQB1_03_01	SARS-CoV:5047_5062 2019-nCoV:5070_5085 MERS-CoV:5056_5071 RaTG13-CoV:5069_5084	4.19	0	0.42
ORF1ab	YFKYWDQTY	HLA-B15:01 HLA-B15:02 HLA-C07:02	2019-nCoV:4678_4686 RaTG13-CoV:4677_4685 SARS-CoV:4655_4663 MERS-CoV:4664_4672	4.08	7.33	0
ORF1ab	MGWDYPKCDR	HLA-A31:01	2019-nCoV:5007_5016 RaTG13-CoV:5006_5015 SARS-CoV:4984_4993 MERS-CoV:4993_5002	4.07	1.21	0

Figures

a

Coronavirus	Gene	Position	Residue	Bepipred	Bepipred2	Chou	Emini	Karplus	Kolasker	Parker	Score
2019-nCov	S	1	M	0	0	0	0	0	0	0	0
2019-nCov	S	2	F	0	0	0	0	0	0	0	0
2019-nCov	S	3	V	0	0	0	0	0	0	0	0
2019-nCov	S	4	F	0	0	0	0	0	1	0	1
2019-nCov	S	5	L	0	0	0	0	0	1	0	1
2019-nCov	S	6	V	0	1	0	0	0	1	0	2
2019-nCov	S	7	L	0	1	0	0	0	1	0	2
2019-nCov	S	8	L	0	1	0	0	0	1	0	2
2019-nCov	S	9	P	0	1	0	0	0	1	0	2
2019-nCov	S	10	L	0	1	0	0	0	1	0	2

b

Coronavirus	Gene	Position	Residue	Bepipred	Bepipred2	Chou	Emini	Karplus	Kolasker	Parker	Score
2019-nCov	S	1	M	0	0	0	0	0	0	0	0
Value:						-3.411					
Threshold:						0.35					

c

2019-nCov	S	10	L	0	1	0	0	0	1	0	2
Software:						<i>Bepipred2,Kolasker</i>					

Figure 1

a

Coronavirus	Gene	HLA	Position	Peptide	Deep BS	Deep IS	Net aff(nM)	Net rank(%)	MHC rank(%)
2019-nCov	S	HLA-A02:01	2_10	FVFLVLLPL	0.94	0.62	91.09	1	0.14
2019-nCov	S	HLA-A02:06	2_10	FVFLVLLPL	0.95	0.71	53.81	0.66	0.13
2019-nCov	S	HLA-A02:10	2_10	FVFLVLLPL	0.95	0.78	247.3	0.51	0.11
2019-nCov	S	HLA-A02:05	2_10	FVFLVLLPL	0.97	0.85	179.02	1.1	0.2
2019-nCov	S	HLA-C03:04	2_10	FVFLVLLPL	0.96	1	82.69	0.32	0.16
2019-nCov	S	HLA-C03:03	2_10	FVFLVLLPL	0.96	1	82.69	0.32	0.11
2019-nCov	S	HLA-C12:02	2_10	FVFLVLLPL	0.8	0.99	283.46	0.41	0.24
2019-nCov	S	HLA-C12:03	2_10	FVFLVLLPL	0.86	1	93.65	0.42	0.2
2019-nCov	S	HLA-A02:01	2_11	FVFLVLLPLV	0.97	0.56	46.44	0.59	0.2
2019-nCov	S	HLA-A02:06	2_11	FVFLVLLPLV	0.98	0.65	23.75	0.3	0.15

b

Coronavirus	Gene	HLA	Position	Peptide	Net_rank(%)	Mix_rank(%)
2019-nCoV	S	DPA1_01_03_DPB1_04_01	1_12	MFVFLVLLPLVS	7	1.09
2019-nCoV	S	DPA1_01_03_DPB1_02_01	35_58	GVYYPDKVFRSSVLHSTQDLFLPF	8.5	1.26
2019-nCoV	S	DPA1_01_03_DPB1_02_01	40_51	DKVFRSSVLHST	9	1.41
2019-nCoV	S	DPA1_01_03_DPB1_02_01	41_52	KVFRSSVLHSTQ	9	0.51
2019-nCoV	S	DPA1_01_03_DPB1_04_01	45_60	SSVLHSTQDLFLPFFS	5	1.04
2019-nCoV	S	DPA1_01_03_DPB1_02_01	53_72	DLFLPFFSNVTWVFHAIHVSG	4.5	0.89
2019-nCoV	S	DPA1_01_03_DPB1_04_01	132_145	EFQFCNDPFLGVVY	4.5	0.61
2019-nCoV	S	DPA1_01_03_DPB1_02_01	132_148	EFQFCNDPFLGVVYHKN	4.5	1.06
2019-nCoV	S	DPA1_01_03_DPB1_02_01	132_149	EFQFCNDPFLGVVYHKNN	6	1.58
2019-nCoV	S	DPA1_01_03_DPB1_02_01	132_150	EFQFCNDPFLGVVYHKNNK	8	0.86

Figure 2

a

Gene	Peptide	B_score	HLA I Frequency(%)	HLA II Frequency(%)	Virus Count
S	YLQPRTFLL	2.33	32.61	0	2
S	IAFNHPIQV	2.22	29.76	0	1
S	MIAAYTAAL	0.89	26.18	0	1
S	YAPEPITSL	4.78	25.45	0	1
S	MIAQYTSAL	1.72	24.83	0	2
S	FVYDAYQNL	2.33	24.8	0	1
S	FSFGVTQEY	2.56	22.83	0	1
S	WTFGAGAAL	1.7	21.33	0	3
S	IAIPTNFTI	1.72	20.5	0	2
S	KQFANGFW	1.78	20.49	0	1

b

Gene	Peptide	B_score	HLA I Frequency(%)	HLA II Frequency(%)	Virus Count
S	YLQPRTFLL	2.33	32.61	0	2
HLA:	HLA-A02:01,HLA-A02:07,HLA-A02:06,HLA-A02:03,HLA-A32:01,HLA-A02:10,HLA-C07:02,HLA-C06:02,HLA-C03:04,HLA-C08:01,HLA-C03:03,HLA-C14:02,HLA-C15:02,HLA-C12:02,HLA-C12:03,HLA-C08:03,HLA-C02:02,HLA-C05:01,HLA-C15:05				

c

Gene	Peptide	B_score	HLA I Frequency(%)	HLA II Frequency(%)	Virus Count
S	YLQPRTFLL	2.33	32.61	0	2
Locate:	2019-nCoV:269_277,RaTG13-CoV:278_286				

Figure 3