 doi:10.20944/preprints202008.0092.v1

COVID-19 Vaccine Candidates by Identification of B and T Cell Multi-Epitopes Against SARS-COV-2

Suresh Kumar*, Sarmilah Mathavan, Wee Jia Jin, Nur Azznira Bt Azman, Devindren Subramanaiam, Nur Afiqah Binti Zainalabidin, Dhivashini Lingadaran, Zainah Binti Abdul Sattar, Danniya Lakshmi Manickam, Priscilla Sheba Anbananthan, Johan Ahmad Taqiyuddin, Yuvapriya Thevarajan

Faculty of Health and Life Sciences, Management and Science University University Drive, Off Persiaran Olahraga, 40100 Shah Alam, Selangor Malaysia

ORCID ID:

Suresh Kumar: https://orcid.org/0000-0001-5682-0938

Sarmilah Mathavan https://orcid.org/0000-0001-6105-502X

Wee Jia Jin: https://orcid.org/0000-0002-2295-0467

Nur Azznira Bt Azman: https://orcid.org/0000-0002-1139-3888

Devindren Subramanaiam: https://orcid.org/0000-0002-0029-4630

Nur Afiqah Binti ZainalAbidin: https://orcid.org/0000-0002-3927-8945

Dhivashini Lingadaran https://orcid.org/0000-0002-9587-9462

Zainah Binti Abdul Sattar https://orcid.org/0000-0003-1978-1102

Danniya Lakshmi Manickam: https://orcid.org/0000-0001-7400-327X

Priscilla Sheba Anbananthan: https://orcid.org/0000-0002-9959-9687

Yuvapriya Thevarajan: https://orcid.org/0000-0001-6162-9898

***Corresponding Author:**

Suresh Kumar, PhD

Senior Lecturer of Bioinformatics

Faculty of Health and Life Sciences, Management and Science University

University Drive, Off Persiaran Olahraga, 40100 Shah Alam, Selangor

Malaysia

Email: sureshkumar@msu.edu.my

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2020 doi:10.20944/preprints202008.0092.v1

Running title: COVID-19 B and T cell epitopes

Abstract:

Coronavirus disease (COVID-19) is a new discovered strain where WHO officially declares the disease as COVID-19 while the virus responsible for it called Severe Acute Respiratory Syndrome Coronavirus 2 or SARS-CoV-2. The incubation period of this disease is between 14 days. Ordinary clinical symptoms that reported around the world include fever, cough, fatigue, diarrhoea and vomiting as well as asymptomatic for certain people. Infection is spread mainly through broad droplets. In early March 2020, WHO again has announced that COVID-19 is a pandemic with currently no specific treatment. The potential use of SARS-COV-2 proteome as a vaccine candidate by analysing through B-cell and T-cell antigenicity by using a immunoinformatics approach as a vaccine development early stage. In this study, we used consensus sequence for SARS-COV-2 proteome that was retrieved from NCBI database. VaxiJen 2.0 was mainly used to identify the antigenic property of SARS-COV-2 proteins. IEDB then used to analyse the B-cell epitope, the presence of T cell immunogenic epitope in SARS-COV-2 proteins was obtained by using compromise method of MHC class I and II tools that accessible respectively using ProPred-1 server and MHC II Binding Prediction in IEDB database. The best epitopes of B and T-cell epitopes were predicted with high antigencity and the information is disseminated through web-based database resource (https://covid-19.omicstutorials.com/epitopes/). This study will be useful to find a new epitope-based candidate for SARS-COV-2. However, further study needs to be done for the next stages of vaccine development.

Keywords: COVID-19, SARS-CoV-2, epitopes, B-cell, T-cell, immuno informatics, MHC-I, MHC-II

Introduction

The causative source of COVID-19 infection is coronavirus 2 (SARS-CoV-2) with extreme severe respiration syndrome. Coronavirus was known to be encased by β-coronavirus, positive single-strand consisting of a large number of RNA viruses that transmit the disease in humans, as well as animals. The SARS-COV-2 proteome consists of the spike (S), envelope (E), nucleocapsid (N), membrane (M), orf3a, or6, orf7a, orf8 and orf10 (Srinivasan et al. n.d.).

Despite low mortality rates, this virus includes high virulence and infectivity. The COVID-19 signs occur after 5.2 days of incubation. Most typical symptoms of COVID-19 disease include

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2020 doi:10.20944/preprints202008.0092.v1

fever, coughing, and exhaustion, although other indications involve sputum, nausea, haemoptysis, diarrhoea, shortness of breath, and lymphopenia. COVID-19 is suspected to invade lung alveolar epithelium as an input channel using the angiotensin-converting enzyme II receptor-mediated endocytosis. Patients hospitalized with COVID-19 reported an increase in leukocyte quantity, abnormal breathing patterns and higher plasma amounts of pro-inflammatory cytokines (Kumar 2020a).

Effective vaccination is important to contain the pandemic outbreak of SARS-COV-2. Vaccine trials are ongoing but vaccine development can take several months to years. Lots of ongoing research concentrate on SARS-COV-2 virus spike protein (Abraham Peele et al. 2020; Bhattacharya et al. 2020; Kumar 2020b, 2020c). The existence of pre-existing memory T cells in humans with the capacity for recognizing SARS-CoV-2 is little understood (Chen and John Wherry 2020; Grifoni et al. 2020). Such knowledge is of urgent significance and will aid in the design of vaccines and facilitate the evaluation of immunogenicity of vaccine candidates. Much of the epitope research focussed on the virus' spike protein and had insufficient knowledge about MHC-I and MHC-II alleles (Naz et al. 2020).

Many studies identified drug and vaccine candidates successfully for various bacteria and virus using reverse vaccinology approach (Kumar 2011, 2015; Kumar and Ramanujam 2020; Omeershffudin and Kumar 2019). In this study, the development of peptide vaccine design has identified by reverse vaccinology, which aims to classify possible candidates for the vaccine considering all the proteome of the virus with the prediction of MHC-I (47 alleles) and MHC-II (27 set alleles) with high antigenicity. The information of all epitopes was developed as a publicly available epitope database (https://covid-19.omicstutorials.com/epitopes/) to bind any class of HLA 1 and HLA 2 all over the SARS-CoV-2 proteome.

Material and Methods

Sequence retrieval

The proteome sequence of novel coronavirus which was submitted at National Centre of Biological Information (NCBI) database from the first report of Wuhan seafood market namely Spike glycoprotein (YP_009724390.1), Nucleocapsid phosphoprotein (YP_009724397.2), membrane glycoprotein (YP_009724393.1), an envelope protein (YP_009724392.1),orf1ab polyprotein (YP_009724389.1), ORF3a protein (YP_009724391.1), ORF6 protein

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2020 doi:10.20944/preprints202008.0092.v1

(YP_009724394.1), ORF7a protein (YP_009724395.1), ORF8 protein (YP_009724396.1), ORF10 protein (YP_009725255.1) (accessed 6 th May 2020) was used to search against betacoronavirus protein database available at NCBI using blastp with default parameters (Grifoni et al. n.d.). The sequence was selected from the same SARS-COV-2 viral sequences which have only 100% coverage and with more than 98% percentage identity from different geographical locations. The sequence which has partial and incomplete sequence are omitted for analysis.

The workbench Virus Pathogen Resource (VIPR) is used for the study of sequence alignment and sequence variation (Pickett et al. 2012). The sequences were matched using ViPR 's MUSCLE algorithm. The sequences which have poor alignment were removed during analysis. For polymorphism calculation MUSCLE is used for multiple sequence alignment. Using VIPR, the consensus protein sequences were obtained and used for further epitope analysis (S, S, and RH 2019).

B-cell epitope prediction

The collected consensus sequence of SARS-COV-2 sequences of surface glycoprotein (S), nucleocapsid phosphoprotein (N), membrane glycoprotein (M), ORF1a and ORF1ab, ORF3a, ORF6, ORF7a, ORF8, and ORF10 were used for B-Cell epitope prediction. The B-Cell epitope prediction was made through Bepipred Linear Epitope Prediction 2.0 (Anon n.d.) available at Immune Epitope Database and Analysis Resources (IEDB) with the setting of 0.55 threshold and the sequence which are more than seven amino acids were considered for prediction.

Prediction of T cell binding epitopes

ProPred I server (Singh and Raghava 2002) was used to predict HLA class I epitopes with default threshold and selected to recognize epitopes that bind to 47 alleles of HLA I. For HLA-class II epitope prediction, MHC-II binding predictions used available in IEDB (Andreatta et al. 2015). The prediction method was selected IEDB recommended with allele selection of 27 full HLA reference set with a default length of 15 with adjusted rank.

Antigenicity analysis

The predicted epitopes were analysed via Vaxijen Server version 2.0 (Doytchinova and Flower 2007) to assess their antigenicity with a threshold of 5 and above for the virus. The epitopes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2020 doi:10.20944/preprints202008.0092.v1

predicted with Vaxijen with 0.5 and above are considered epitopes which have high antigenicity.

Result

Data retrieval

A total of spike glycoprotein (100 sequences), nucleoprotein (85 sequences), membrane protein (43 sequences), an envelope protein (15 sequences), orflab (99 sequences), orf3a (75 sequences), orf6 (27 sequences), orf7a (20 sequences), orf8 (22 sequences), orf10 (6 sequences) belonging to different geophraphical locations were downloaded from the Severe acute respiratory syndrome coronavirus 2 data hub from NCBI. The consensus sequence obtained from multiple alignment are used for epitope prediction. The consensus sequence used for prediction are available at: https://covid-19.omicstutorials.com/epitopes/.

B-cell epitope prediction

Based on the prediction of linear B-cell epitopes from the IEDB server, 12 epitopes for Spike Protein, 5 epitopes for Nucleoprotein, 1 epitope for a membrane protein, 1 epitope for envelope protein, 81 epitopes for orflab, 2 epitopes for orf3a, 1 epitope for orf6 and 1 epitope for orf7a based on antigenic propensity prediction by VaxiJen server analysis..

T-cell epitope prediction

We identified 92 epitopes for spike protein, 6 epitopes for nucleoprotein, 36 epitopes for a membrane protein, 6 protein for envelope protein, 68 epitopes for orlab, 5 epitopes of orf3a, 16 epitopes for orf6, 15 epitopes for orf7a, 44 epitopes for orf8, 17 epitopes for orf10 predicted for MHC-I which antigenicity score 5 and above. For MHC-II epitopes, 560 for spike protein, 203 for nucleocapsid, 106 for membrane protein, 35 for envelope protein, 129 for orlfab, 101 for orf3a, 18 for orf6, 52 for orf7a, 55 for orf8, 135 for orf10 with antigenicity 5 and above based on VaxiJen antigenicity score.

Discussion

In our study, we used all structural, non-structural and accessory proteins of SARS-COV-2 for B and T-cell epitope prediction using various computational tools through immunoinformatics

approach (Patronov and Doytchinova 2013; Raoufi et al. 2020; Tomar and De 2010). Also, we used all 47 alleles for MHC-I and 27 alleles for MHC-II binding prediction. This identification will be crucial for vaccine design against COVID-19. In B-cell epitope prediction (Bettencourt et al. 2020; Prachar et al. 2020; Rock, Reits, and Neefjes 2016), 12 epitopes for Spike Protein, 5 epitopes for Nucleoprotein, 1 epitope for a membrane protein, 1 epitope for envelope protein, 81 epitopes for orflab, 2 epitopes for orf3a, 1 epitope for orf6 and 1 epitope for orf7a predicted based on antigenicity prediction based on threshold 0.5 and above. We discarded the epitopes which have a non-antigenicity characteristic.

Similarly, for T-cell epitope prediction of MHC-I (Smith et al. 2020), we identified 92 epitopes for spike protein, 6 epitopes for nucleoprotein, 36 epitopes for a membrane protein, 6 protein for envelope protein, 68 epitopes for orlab, 5 epitopes of orf3a, 16 epitopes for orf6, 15 epitopes for orf7a, 44 epitopes for orf8, 17 epitopes for orf10 predicted for MHC-I which antigenicity score 5 and above and discarded which have the non-antigenicity characteristic. We selected only best epitopes which have higher antigenicity score for database resource construction. The evaluated epitopes can be used against SARS-COV-2 as a multi-epitope vaccine candidate for COVID-19.

Database availability:

The SARS-COV-2 protein sequence of spike (S), envelope (E), nucleocapsid (N), membrane (M), orf3a, or6, orf7a, orf8 and orf10 predicted epitopes of B-cell, MHC-I with 47 alleles binding prediction and MHC-II with 27 alleles binding prediction of T-cell epitopes were publicly available through the resources: https://covid-19.omicstutorials.com/epitopes/

Conclusion

The epitope prediction evaluated both B and T cell epitopes for all SARS-COV-2 proteome for high antigenicity epitopes. The epitopes of MHC-I with 47 allele binding and MHC-II with 27 allele binding prediction will help to understand the mechanism of pathogenesis. However, the designed epitopes need further wet-lab validation.

Competing Interest Statement: The authors have declared no competing interest.

Funding Statement: There was no funding received for the completion of this study.

 doi:10.20944/preprints202008.0092.v1

References

Abraham Peele, K., T. Srihansa, S. Krupanidhi, Vijaya Sai Ayyagari, and T. C. Venkateswarulu. 2020. "Design of Multi-Epitope Vaccine Candidate against SARS-CoV-2: A *in-Silico* Study." *Journal of Biomolecular Structure and Dynamics* 1–9.

Andreatta, Massimo, Edita Karosiene, Michael Rasmussen, Anette Stryhn, Søren Buus, and Morten Nielsen. 2015. "Accurate Pan-Specific Prediction of Peptide-MHC Class II Binding Affinity with Improved Binding Core Identification." *Immunogenetics* 67(11–12):641–50.

Anon. n.d. "BepiPred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes." Retrieved August 2, 2020 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5570230/).

Bettencourt, Paulo, Julius Müller, Annalisa Nicastri, Daire Cantillon, Meera Madhavan, Philip D. Charles, Carine B. Fotso, Rachel Wittenberg, Naomi Bull, Nawamin Pinpathomrat, Simon J. Waddell, Elena Stylianou, Adrian V. S. Hill, Nicola Ternette, and Helen McShane. 2020. "Identification of Antigens Presented by MHC for Vaccines against Tuberculosis." *Npj Vaccines* 5(1):1–14.

Bhattacharya, Manojit, Ashish R. Sharma, Prasanta Patra, Pratik Ghosh, Garima Sharma, Bidhan C. Patra, Sang Soo Lee, and Chiranjib Chakraborty. 2020. "Development of Epitope-Based Peptide Vaccine against Novel Coronavirus 2019 (SARS-COV-2): Immunoinformatics Approach." *Journal of Medical Virology* 92(6):618–31.

Chen, Zeyu and E. John Wherry. 2020. "T Cell Responses in Patients with COVID-19." *Nature Reviews Immunology* 1–8.

Doytchinova, Irini A. and Darren R. Flower. 2007. "VaxiJen: A Server for Prediction of Protective Antigens, Tumour Antigens and Subunit Vaccines." *BMC Bioinformatics*.

Grifoni, A., J. Sidney, Y. Zhang, … RH Scheuermann-Cell host &., and undefined 2020. n.d. "A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2." *Elsevier*.

Grifoni, Alba, Daniela Weiskopf, Sydney I. Ramirez, Jose Mateus, Jennifer M. Dan, Carolyn Rydyznski Moderbacher, Stephen A. Rawlings, Aaron Sutherland, Lakshmanane Premkumar, Ramesh S. Jadi, Daniel Marrama, Aravinda M. de Silva, April Frazier, Aaron F. Carlin, Jason A. Greenbaum, Bjoern Peters, Florian Krammer, Davey M. Smith, Shane

Crotty, and Alessandro Sette. 2020. "Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals." *Cell* 181(7):1489.

Kumar, Suresh. 2011. "Molecular Modeling and Identification of Substrate Binding Site of Orphan Human Cytochrome P450 4F22." *Bioinformation* 7(4):207–10.

Kumar, Suresh. 2015. "Computational Identification and Binding Analysis of Orphan Human Cytochrome P450 4X1 Enzyme with Substrates." *BMC Research Notes* 8(1):9.

Kumar, Suresh. 2020a. "COVID-19: A Drug Repurposing and Biomarker Identification by Using Comprehensive Gene-Disease Associations through Protein-Protein Interaction Network Analysis."

Kumar, Suresh. 2020b. "Drug and Vaccine Design against Novel Coronavirus (2019-NCoV) Spike Protein through Computational Approach." *Preprints (Www.Preprints.Org)* (February).

Kumar, Suresh. 2020c. "Drug and Vaccine Design against Novel Coronavirus (2019-NCoV) Spike Protein through Computational Approach."

Kumar, Suresh and Meera Ramanujam. 2020. *Computational Prediction of Novel Broad-Spectrum Drug Targets Against Vibrio Cholerae by Integrated Genomics and Proteomics Approach*. Vol. 16.

Naz, Anam, Fatima Shahid, Tariq Tahir Butt, Faryal Mehwish Awan, Amjad Ali, and Arif Malik. 2020. "Designing Multi-Epitope Vaccines to Combat Emerging Coronavirus Disease 2019 (COVID-19) by Employing Immuno-Informatics Approach." *Frontiers in Immunology* 11(July):1–13.

Omeershffudin, Umairah Natasya Mohd and Suresh Kumar. 2019. "In Silico Approach for Mining of Potential Drug Targets from Hypothetical Proteins of Bacterial Proteome." *International Journal of Molecular Biology* 4(4):145–52.

Patronov, Atanas and Irini Doytchinova. 2013. "T-Cell Epitope Vaccine Design by Immunoinformatics." *Open Biology* 3(JAN).

Pickett, Brett E., Douglas S. Greer, Yun Zhang, Lucy Stewart, Liwei Zhou, Guangyu Sun, Zhiping Gu, Sanjeev Kumar, Sam Zaremba, Christopher N. Larsen, Wei Jen, Edward B. Klem, and Richard H. Scheuermann. 2012. "Virus Pathogen Database and Analysis Resource (ViPR): A Comprehensive Bioinformatics Database and Analysis Resource for

the Coronavirus Research Community." *Viruses* 4(11):3209–26.

Prachar, Marek, Sune Justesen, Daniel Bisgaard Steen-Jensen, Stephan Thorgrimsen, Erik Jurgons, Ole Winther, and Frederik Otzen Bagger. 2020. "COVID-19 Vaccine Candidates: Prediction and Validation of 174 SARS-CoV-2 Epitopes." *BioRxiv* 2020.03.20.000794.

Raoufi, Ehsan, Maryam Hemmati, Samane Eftekhari, Kamal Khaksaran, Zahra Mahmodi, Mohammad M. Farajollahi, and Monireh Mohsenzadegan. 2020. "Epitope Prediction by Novel Immunoinformatics Approach: A State-of-the-Art Review." *International Journal of Peptide Research and Therapeutics* 26(2):1155–63.

Rock, Kenneth L., Eric Reits, and Jacques Neefjes. 2016. "Present Yourself! By MHC Class I and MHC Class II Molecules." *Trends in Immunology* 37(11):724–37.

S, Phadke, Macherla S, and Scheuermann RH. 2019. "Database and Analytical Resources for Viral Research Community." *Reference Module in Life Sciences*.

Singh, Harpreet and G. P. S. Raghava. 2002. "ProPred: Prediction of HLA-DR Binding Sites." *Bioinformatics*.

Smith, Christof C., Sarah Entwistle, Caryn Willis, Steven Vensko, Wolfgang Beck, Jason Garness, Maria Sambade, Eric Routh, Kelly Olsen, Julia Kodysh, Timothy O'Donnell, Carsten Haber, Kirsten Heiss, Volker Stadler, Erik Garrison, Oliver C. Grant, Robert Woods, Mark Heise, Benjamin G. Vincent, and Alexander Rubinsteyn. 2020. "Landscape and Selection of Vaccine Epitopes in SARS-CoV-2." *BioRxiv* 2020.06.04.135004.

Srinivasan, Suhas, Hongzhu Cui, Ziyang Gao, Ming Liu, Senbao Lu, Winnie Mkandawire, Oleksandr Narykov, Mo Sun, and Dmitry Korkin. n.d. "Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins." *Mdpi.Com*.

Tomar, Namrata and Rajat K. De. 2010. "Immunoinformatics: An Integrated Scenario." *Immunology* 131(2):153–68.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2020 doi:10.20944/preprints202008.0092.v1

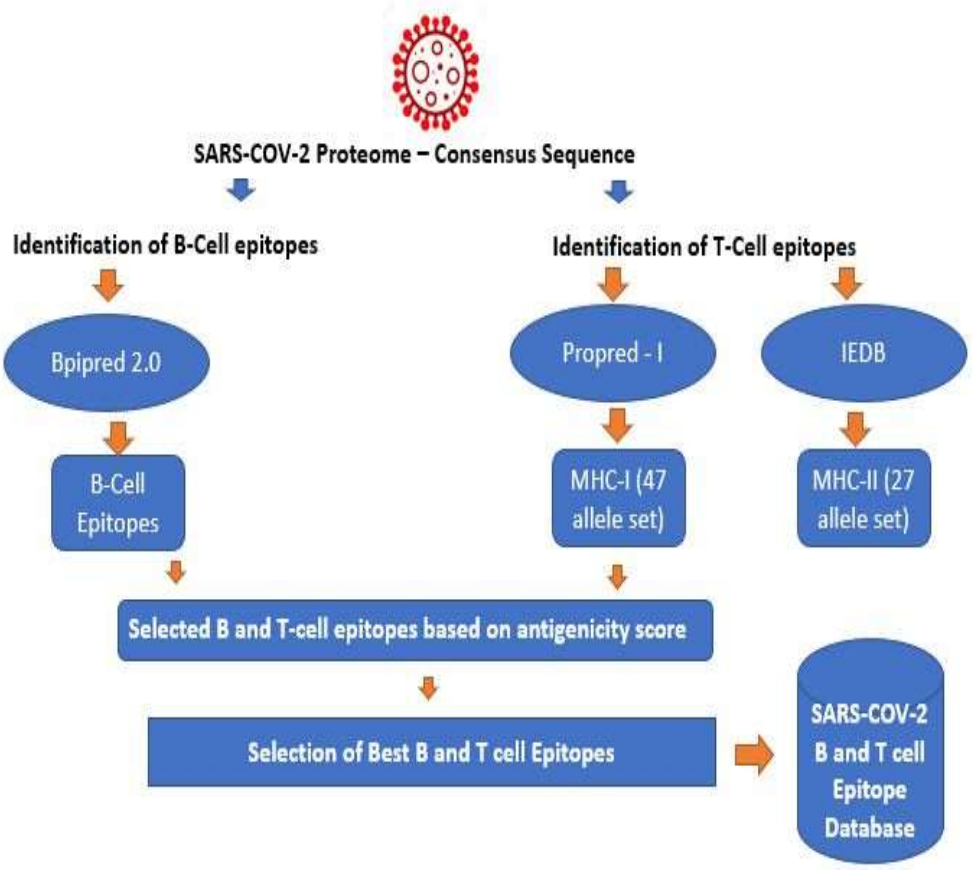

Figure 1. Reverse vaccinology approaches used to identify potential B and T-cell epitopes from the proteome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2020 doi:10.20944/preprints202008.0092.v1

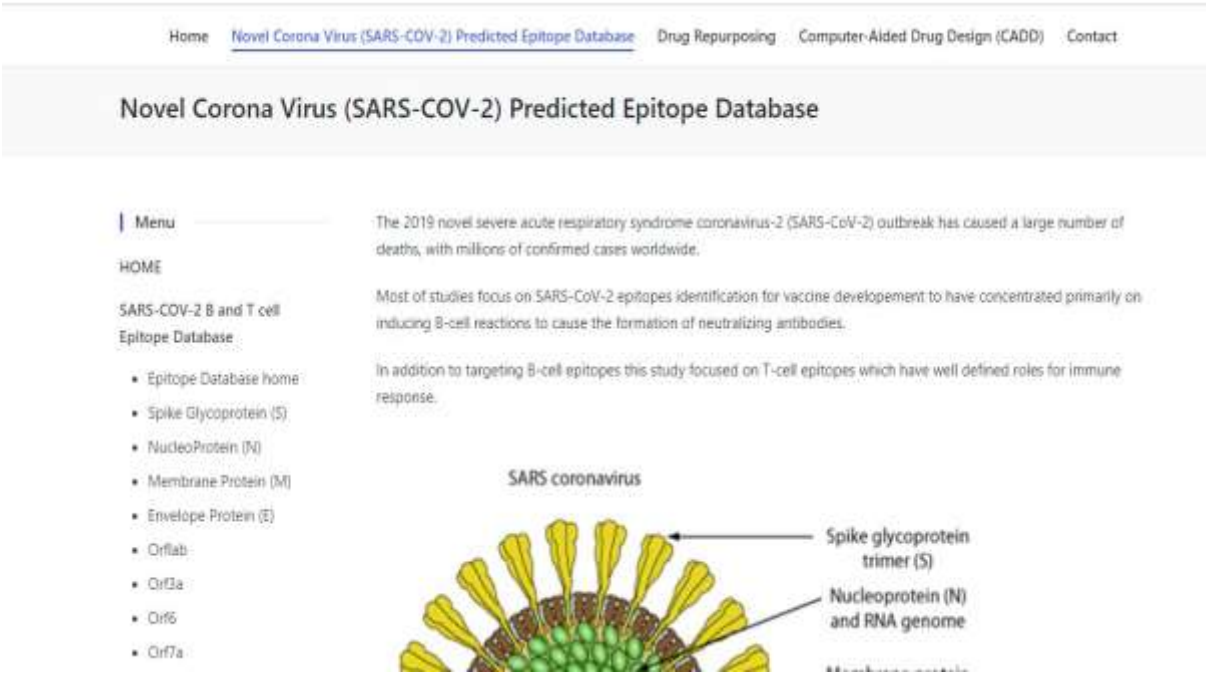

Figure 2: The homepage of SARS-COV-2 predicted epitope database available at: https://covid-19.omicstutorials.com/

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 August 2020 doi:10.20944/preprints202008.0092.v1

Table 1: Predicted potential B-cell , MHC-I and MHC-II T-cell epitopes of SARS-COV-2

	B Cell Epitope	MHC-I	MHC-II
Spike Glycoprotein (S)	12	92	560
NucleoProtein (N)	5	6	203
Membrane Protein (M)	1	36	106
Envelope protein (E)	1	6	35
Orflab	81	68	129
Orf3a	2	5	101
Orf6	1	16	18
Orf7a	1	15	52
Orf8	1	44	55
Orf10	1	17	135