Supplementary File 1

Title: Does immune recognition of SARS-CoV2 epitopes vary between different ethnic groups?

Tungadri Bose[$]*, Namrata Pant[$], Nishal Kumar Pinna, Subhrajit Bhar, Anirban Dutta*, Sharmila S. Mande*

TCS Research, Tata Consultancy Services Limited, Pune, India

[$]Equal contributors
*Corresponding authors

Identification of viral peptides as potential vaccine candidates

Since the temporal changes in the SARS-CoV2 genome did not seemed to appreciably change the epitope recognition ability of the host, we further investigated if a minimal set of antigenic peptides could be identified which may be utilized to apriori sensitize the human immune system, thereby providing resistance to Covid-19 infection at a global scale. Based on the 487 epitopes (and their variants) that were recognized (with a predictions scores ≥ 0.95) by at least 5% of the individuals in each of the ethnicities, Euler plots (Supplementary Figure 10) for identifying the minimal set of SARS-CoV2 epitopes (both CD8 and CD4) were generated. A total of 11 CD8 and 17 CD4-specific epitopes were found to match our criteria (Table 2, Supplementary Table 10). Of these, 8 CD8 and 10 CD4-specific epitopes were also confirmed as potential vaccine subunits by the software tool VaxiJen (1) with a threshold ≥ 0.5. Further, several of the identified epitopes were found to map (with 100% identity and 100% overlap) with the SARS proteome, thereby indicating at the possibility of using these epitopes to target multiple lineage B *Betacoronavirus* strains (Table 2). However, the overlap of these epitopes to the MERS proteome was extremely limited. The list of SARS and MERS protein sequences used in this study have been listed in Supplementary Table 11. In addition to the above-mentioned predicted epitopes, 26 CD8-specific epitopes were also found, which can be explored further as potential generic vaccine candidates, except for East Asian

(EAS) ethnicities (Supplementary Table 10). Among these EAS ethnicities 31 additional CD8-specific potential vaccine candidates were found. These were unique to EAS and could possibly provide adequate protection to individuals of this super-population. Similar patterns among the CD4-specific epitopes were also noted (Supplementary Table 10). Overall, results indicated that sensitizing the immune system with any of these (or a cocktail of these) predicted epitopes mentioned in Supplementary Table 10 could potentially provide protection to Covid-19 infection at a global scale.

Comparative analysis of number of variations in epitopes and the corresponding proteins in SARS-CoV2

Only 25 out of 4194 variants of 'reference' epitopes (VREs) were observed to occur more frequently (frequency ≥ 0.5%) among the studied genomes (MFVREs). While it was anticipated that the effect of these MFVREs will not be significant in determining the antigen response to SARS-CoV2, it was interesting to explore if there were any epitopes which might experience a selection pressure and accumulate a larger proportion of VREs. In such a scenario, the less frequent VREs (LFVREs) could also play a role in determining the overall response to Covid-19 at a population level. An analysis was therefore performed (Supplementary Fig. 9) wherein the variations accumulated over the length of the epitopes were compared to the variations picked up by the proteins (harboring the antigenic peptide) for all the 40,342 genomes. The total number of variations reported across the studied genomes were normalized by the length of the epitope (and proteins) and plotted in Supplementary Fig. 9. Results indicate that the median of the (normalized) number of variations observed in the epitopes for each protein were comparable to the number of amino acid changes on the protein with respect to the reference genome. We conclude that there may not be any selection pressure on the epitopic regions of the SARS-CoV2 proteins which might have otherwise aided the pathogen in immune evasion.

REFERENCES:

1. Doytchinova IA, Flower DR. 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. 1. BMC Bioinformatics 8:1–7.