Designing of cytotoxic and helper T cell epitope map provides insights into the highly contagious nature of the pandemic novel coronavirus SARS-CoV2

Seema Mishra

Department of Biochemistry

School of Life Sciences

University of Hyderabad, India

Email: seema_uoh@yahoo.com

Abstract:

Novel coronavirus, SARS-CoV2, has emerged one of the deadliest pathogens of this century creating a pandemic. Belonging to betacoronavirus family, it spreads through human contact and even through asymptomatic transmission. Till date, there is no known treatment in the form of drugs or vaccines, despite several attempts since it emerged. Several vaccines are in pre-clinical and two in clinical trials as of 4th April 2020 as per WHO document. Here, in order to develop subunit vaccine, attempts have been made to find globally conserved epitopes from all ten SARS-CoV2 proteins as there is no clear information on the virulence of these proteins. Using computational tools, a ranked list of probable immunogenic, promiscuous epitopes generated through all three main stages of antigen processing and presentation pathway has been generated. Moreover, on the way to finding these epitopes, several useful insights were gleaned. One of the most important insights is that all of the proteins in this pathogen present unique epitopes, so that if one protein's function is hindered by the immune system,

other proteins can continue to assist in virus survival. Due to presence of these unique epitopes in all SARS-CoV2 proteins, a stronger immune response generated may lead to immunopathology and consequently, less chances of human survival. These epitopes, after due validation *in vitro*, may thus need to be presented to the human body in that form of subunit vaccine that avoids such immunopathologies.

Introduction

Novel coronavirus (SARS-CoV2), also known as 2019-nCoV, causes Covid19 disease with significant mortality rate. As there is currently no known cure, vaccine design and development is urgently required. Despite 77 drugs against viral spike protein being identified by world's fastest supercomputer, Summit, (1), Immunoinformatics tools will prove crucial (2, 3). As of 4th April 2020, WHO has put forward a draft which identifies 2 vaccines in clinical evaluation and 60 candidate vaccines in preclinical evaluation (4). This study presents several such novel cytotoxic and helper T cell epitopes against ORF1ab protein and helper T cell epitopes against all other proteins.

SARS-CoV2 genome submitted by CDC, Atlanta (GenBank accession number: MT106054.1 submitted on 24-Feb-2020) is 29882 bp in length. Being 100% identical to reference sequence, NC_045512.2 from Wuhan, China, it harbors multiple structural, non-structural and accessory proteins essential or playing a role at various stages of the viral life cycle. This SARS-CoV2 genome is found 82.3% identical to SARS-CoV genome (NC_004718.3), using NCBI BLASTn tool. T cell epitopes against several proteins in SARS and MERS species have been identified (5, 6). In brief, the sequence of proteins in its RNA genome as per this GenBank accession information is as follows: 5′-ORF1ab-S

(Spike/Surface)–ORF3a-E (Envelope)-M (Membrane)-ORF6-ORF7a-ORF8-N (Nucleocapsid)-ORF10-polyA tail-3', which are usually seen in betacoronaviruses (7). ORF1ab, a polyprotein, encodes several non-structural proteins, 15 in number identified in this genome sequence annotation, including RNA-dependent RNA polymerase (RdRP). The role of structural proteins is determined from their homology to SARS-CoV as well as few experiments (8). Expression, localization and function of some SARS-CoV2 accessory proteins is as yet unclear, although several such proteins have been characterized in SARS-CoV (9) and the roles may be similar in the two viruses. Sequencing studies suggest that the most abundant transcript was N RNA followed by S, ORF7a, ORF3a, ORF8, M, E, ORF6, and ORF7b (10; ORF7b is identified in this paper). In view of scarcity of data on relevance and roles of these proteins, any one or more of these proteins may act as prime vaccine candidates. Hence, all of these proteins were used for T cell epitope prediction for the purpose of peptide-based subunit vaccine design and further analyses. The fact that this approach may be better also arises from previous studies on related SARS-CoV virus (11), wherein more than 50% of the patients had T cell responses against at least one of the two proteins tested, and 25% showed responses against both proteins.

The advantages of peptide subunit vaccine as opposed to DNA and live attenuated virus vaccine is that they do not contain live components and so are considered safe. Moreover, they present an antigen or a set of antigens to the immune system with lower risk of side effects (12). These are also applicable to people with weakened immune response, which the old people have and are, therefore, prime targets in this SARS-CoV2 infection. Henec, this study has been done with the objectives of finding novel CTL and HTL epitopes and helping glean many important insights along the way.

Results and Discussion:

Cytotoxic T lymphocyte (CTL) epitopes

Two prediction algorithms were used to generate a consensus list of nonameric CTL epitopes. The consensus list was chosen to increase prediction accuracy from two different algorithms. While

NetCTLpan uses neural network algorithm, PickPocket works on the basis of position-specific weight matrices. NetCTLpan, in addition to HLA binding, also predicts TAP-transporter binding and C-terminal proteasome cleavage predictions. Total number of CTL epitopes generated was 9621 across ten SARS-CoV2 proteins including ORF1ab polyprotein. A total of 122 epitopes were enlisted. These common, promiscuous CTL epitopes are enlisted in Table 1 as ranked order for ORF1ab. For other proteins, it may be observed/enlisted from (13). It is seen that out of a few common promiscuous epitopes for surface protein across prediction algorithms (13), one of these epitopes, FVFLVLLPL, signal peptide in surface/spike protein, has been found to harbor a mutation, L5F, in many strains of 13 countries in distinct phylogenetic clades and L8V/W mutation is present in Hong Kong (14). These authors further suggest that L5F mutation might be a sequencing artifact. The highest number of common top-ranking epitopes is seen in the case of nsp7 of ORF1ab followed by ORF10, ORF8, ORF6 and ORF3a proteins. Among structural proteins, envelope protein provided the highest number of such epitopes. Venn diagram analysis showed no common epitopes at all across proteins and alleles. Even though SARS-CoV2 RBD (331-527) is shown to harbor epitopes for eliciting neutralizing antibodies (14, 15), this region is not present in this data for CTL epitopes. However, receptor-binding motif (RBM) region (437-508), the ACE-2 binding motif of this RBD provided immunogenic HTL epitopes which are detailed below in section on promiscuous HTL epitopes. Immunogenicity prediction of these proteins (Table 2) showed that 71 of these 122 epitopes had a positive immunogenicity score. Further, conserved residues between SARS-CoV2 and other HCoV and MERS species were found from multiple sequence alignments and found in several of these epitopes (Supplementary fig. S1). As the NCBI RefSeq sequence of SARS-CoV was unclear in proper annotations for respective proteins, it could not be used in MSA studies. It is observed that most of the epitopes with conserved residues belonged to ORF1ab region (table 2), and epitopes belonging to this region may act as vaccine candidates targeting MERS and other HCoV species, in addition to SARS-CoV2.

Table 1: Top ranked sequences of CTL epitopes common across HLA supertypes (HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39:01, HLA-B*40:01, HLA-B*58:01, HLA-B*15:01) and across the two prediction algorithms used for SARS-CoV2 ORF1ab polyprotein.

Leader protein (nsp1)	nsp2	nsp3	nsp4	3C-like Proteinase	nsp6	nsp7	nsp8	nsp9	nsp10	RdRP	Helicase	3'-5'exonuclease	EndoRNAse	Ribose methyltransferase
HVGEIPVAY	NMMVTNNTF	LVAEWFLAY (not among top scorers in many alleles)	VVAFNTLLF VVAFNTLLF all alleles except HLA-A*02:01, HLA-A*03:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39:01, HLA-B*40:01	QTFSVLACY	MLVYCFLGY all alleles except HLA-A*02:01, HLA-A*24:02, HLA-*07:02, HLA-B*39:01, HLA-B*58:01,	MVSLLSVLL	SSLPSYAAF	CTDDNALAY	FAVDAAKAY	MVMCGGSLY among top scoring in all allelles, except HLA-A*02:01, HLA-A*24:02, HLA-B*08:01, HLA-B*39:01,	YVFCTVNAL all except HLA-A*01:01, HLA-A*03:01, HLA-A*24:02, HLA-B*58:01	HSIGFDYVY present in HLA-A*01:01, HLA-A*03:01, HLA-A*26:01, HLA-B*07:02, HLA-B*27:05, HLA-B*40:01, HLA-B*58:01, HLA-B*15:01;	TTLPVNVAF- all alleles,	YVMHANYIF- all alleles except HLA-A*03:01, HLA-B*27:05
VMVELVAEL				RTILGSALL	LIISVTSNY all alleles except HLA-A*02:01, HLA-A*24:02, HLA-B*08:01, HLA-B*39:01,	SLLSVLLSM	ISMDNSPNL	KSDGTGTIY	STVLSFCAF	TMADLVYAL among top scoring in all allelles, except HLA-A*01:01, HLA-A*03:01, HLA-A*24:02, HLA-B*08:01, HLA-B*27:05, HLA-B*58:01,	FAIGLALYY (top-scorer in few alleles in HLA-A*01:01, HLA-A*26:01, HLA-B*58:01, HLA-B*15:01)		VSIINNTVY- all except HLA-A*02:01	YSLFDMSKF- all except HLA-A*02:01, HLA-B*08:01, HLA-B*39:01
HVQLSLPVL				VSFCYMHHM	FLARGIVFM all alleles except	KMVSLLSVL	AMQTMLFTM	GTGTIYTEL	PANSTVLSF	LMIERFVSL all alleles except A*01:01,			LLLDDFVEI- all except HLA-A*03:01	

					HLA-A*01:01, HLA-A*03:01, HLA-A*24:02, HLA-B*27:05, HLA-B*40:01, HLA-B*58:01,					A*03:01, A*26:01, HLA-B*58:01,			and HLA-B*27:05	
HLKDGTCGL				FLNRFTTTL		CTSVVLLSV	TTFTYASAL	TELEPPCRF	FGGASCCLY				SQLGGLHLL-all except HLA-A*01:01, HLA-A*03:01, HLA-A*26:01 , HLA-B*07:02,	
PQLEQPYVF				HSMQNCVLK all alleles except HLA-A*01:01, HLA-A*02:01, HLA-A*24:02, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39:01, HLA-B*40:01, HLA-B*58:01, HLA-B*15:01		TSVVLLSVL		YFIKGLNNL	ITVTPEANM					
QLEQPYVFI						KLWAQCVQL		ALAYYNTTK	VLSFCAFAV					
RTAPHGHVM						VLLSVLQQL		GMVLGSLAA	YLASGGQPI					
						EMLDNRATL		FVLALLSDL (in some, after around top 30)						
						EAFEKMVSL								

						DVKCTSVVL								
						LAKDTTEAF								
						LHNDILLAK								
						LSMQGAVDI								
						VQLHNDILL								

Table 2: Immunogenic CTL epitopes across proteins, sorted by high HLA-I binding, high immunogenicity, and conservation of residues in multiple sequence alignment (MSA)**;** Epitopes in red are in consensus sequence with HLA-II alleles, blue highlights are for the presence of conserved sequences

Epitope	Protein	Peptide start	Peptide end	Immunogenicity score	Clustering with HLA-II epitopes	Conservation in MSA
FLFLTWICL	Membrane	26	34	0.35397	Singleton	F, L, F semi-conserved
VFAFPFTIY	ORF10	6	14	0.34042	In 10- membered group	No conservation
GIIITVAAF	ORF8	8	16	0.30966	In 5- membered group	I,I semi conserved
IQYIDIGNY	ORF8	71	79	0.30442	Singleton	Fully conserved residues across bat, mers and ncov, I, Q, I in IQYI
NVFAFPFTI	ORF10	5	13	0.30241	In 10- membered group	No conservation
FLAFVVFLL	Envelope	20	28	0.30188	Singleton	last V last L semiconserved
TIAEILLII	ORF6	10	18	0.30101	In 23- membered group	No conservation
FLIVAAIVF	ORF7a	101	109	0.29611	Singleton	No conservation
KVSIWNLDY	ORF6	23	31	0.29343	In 23- membered group	No conservation
VTIAEILLI	ORF6	9	17	0.28951	In 23- membered group	No conservation
MGYINVFAF	ORF10	1	9	0.28694	In 10- membered group	No conservation
YINVFAFPF	ORF10	3	11	0.28259	In 10- membered group	No conservation
SFYEDFLEY	ORF8	103	111	0.28049	Singleton	semi conserved S F E D and conserved L
WNLDYIINL	ORF6	27	35	0.24894	In 23- membered group	No conservation
NLDYIINLI	ORF6	28	36	0.24642	In 23- membered group	No conservation
NTASWFTAL	Nucleocapsi	48	56	0.22775	Singleton	S fully,conserved, F, L sem
TLAILTALR	Envelope	30	38	0.1989	In 8- membered group	L A I L R semi
ILFLALITL	ORF7a	4	12	0.1895	In 6- membered group	No conservation
WLIVGVALL	ORF3a	45	53	0.18314	Singleton	I, V and last L semi conserve
EYHDVRVVL	ORF8	110	118	0.1807	Singleton	No conservation
QVDVVNFNL	ORF10	29	37	0.17787	In 3- membered group	No conservation
AFPFTIYSL	ORF10	8	16	0.1775	In 10- membered group	No conservation
KIILFLALI	ORF7a	2	10	0.16214	In 6- membered group	No conservation
ILLIIMRTF	ORF6	14	22	0.16098	In 23- membered group	No conservation
CVRGTTVLL	ORF7a	23	31	0.1536	Singleton	No conservation
SIWNLDYII	ORF6	25	33	0.15011	In 23- membered group	No conservation
YLYALVYFL	ORF3a	107	115	0.13151	In 7- membered group	No conservation
YLQPRTFLL	Surface/spik	269	277	0.1305	Singleton	First L fully conserved, last L, L semi conserved
LLYDANYFL	ORF3a	139	147	0.11841	In 7- membered group	No conservation
ITLATCELY	ORF7a	23	31	0.10084	Singleton	No conservation
HLVDFQVTI	ORF6	3	11	0.0982	In 6- membered group	No conservation
NSRNYIAQV	ORF10	22	30	0.09731	In 9- membered group	No conservation
IAQVDVVNF	ORF10	27	35	0.09546	In 3- membered group	No conservation
MFHLVDFQV	ORF6	1	9	0.09154	In 6- membered group	No conservation
YFIASFRLF	Membrane	95	103	0.06887	In 14- membered group	Y, F, S, R, L fully conserved, F, F semi-conserved
FPFTIYSLL	ORF10	9	17	0.05708	In 10- membered group	No conservation
FVFLVLLPL	Surface/spik	2	10	0.04076	Singleton	F V F L V SEMI CONSERVE
ELYSPIFLI	ORF7a	95	103	0.03913	Singleton	No conservation
FLYLYALVY	ORF3a	105	113	0.03563	Singleton	No conservation
LTALRLCAY	Envelope	34	42	0.01886	In 8- membered group	first L semi, R semi, LC fully

ORF1ab

Epitope	Protein	From	To	Immunogenicity score	Clustering	Conservation in MSA
LVAEWFLAY	nsp3	1505	1513	0.45285	Singleton	L, L, A semi conserved,
FLARGIVFM	nsp6	184	192	0.3263	Singleton	L,R semi-conserved
HVGEIPVAY	Leader	110	118	0.28861	Singleton	No conserved residue
GTGTIYTEL	nsp9	61	69	0.26744	Singleton	second G and I semi co
FLNRFTTTL	3C-like prot	219	227	0.25596	Singleton	F,L,N semi conserved
LLLDDFVEI	EndoRNAse	297	305	0.24386	Singleton	Second L, D,D,F,V fully
LMIERFVSL	RdRp	854	862	0.24273	is part of 6-membered group	L, M, I semi, E,R fully, F
VMVELVAEL	Leader	84	92	0.23373	Singleton	No conserved residue
HSIGFDYVY	3'-5'exonucl	229	237	0.23318	Singleton	H fully, S semi, D, Y full
VSIINNTVY	EndoRNAse	24	32	0.22161	Singleton	S, first I, N,N,T,V semi-c
KSDGTGTIY	nsp9	58	66	0.22152	Singleton	S, D, second G, I semi
VLSFCAFAV	nsp10	13	21	0.17009	is part of 14-membered group	V semi, L fully, S semi,
ITVTPEANM	nsp10	55	63	0.16515	Singleton	I fully, T,T, E, A, N sem
LHNDILLAK	nsp7	35	43	0.15288	is part of 20-membered group	H,N,I fully, D semi
VQLHNDILL	nsp7	33	41	0.14937	is part of 20-membered group	H,N,I fully, D semi
VVAFNTLLF	nsp4	314	322	0.1449	is part of 16-membered group	second V, T, L semi
LAKDTTEAF	nsp7	41	49	0.13402	is part of 20-membered group	D, A fully
EMLDNRATL	nsp7	74	82	0.11684	Singleton	E,D,A semi, last L fully
RTAPHGHVM	Leader	77	85	0.11636	Singleton	No conserved residue
FAIGLALYY	Helicase	291	299	0.09181	is part of 10-membered group	I,G, last Y fully, A,L,A,L,
TMADLVYAL	RdRp	123	131	0.08282	Singleton	T,M,D,first L, A, last L f
YVMHANYIF	Ribose meth	222	230	0.0822	Singleton	H,A,N,Y,F fully, M,I sem
MLVYCFLGY	nsp6	211	219	0.07782	Singleton	First L, Y,G fully, M, las
YVFCTVNAL	Helicase	355	363	0.07781	Singleton	Y, F, T, N, A, L fully, V,
TTLPVNVAF	EndoRNAse	47	55	0.07705	Singleton	First T, P, N, A fully-cor
SQLGGLHLL	EndoRNAse	243	251	0.07388	Is part of 8-membered group	First L semi-conserved,
CTDDNALAY	nsp9	23	31	0.07355	Singleton	T, A semi
TELEPPCRF	nsp9	67	75	0.06065	Singleton	E, L, P, P, C, f fully, sec
ALAYYNTTK	nsp9	28	36	0.05473	is part of 6-membered group	Y fully, second A, N se
NMMVTNNTF	nsp2	625	633	0.03347	is part of 6-membered group	F semi
QLEQPYVFI	Leader	63	71	0.0049	is part of 2-membered group	Q, I semi

During these CTL epitope identification studies, it was also found that many epitopes same as SARS-CoV epitopes found previously in spike, membrane, nucleocapsid and ORF3a proteins (16) were in the lower ranking positions, in the case of different alleles, and many were not common across epitopes, so confidence could not be gathered in enlisting these. However, in ORF3a case, one epitope harbouring both CD8+ and CD4+ T cell epitopes, PLQASLPFGWLVIGV, among the 3 most frequently recognized by T cells (17) was also present among top-ranked ones in our study (Table 1a). Purely for the sake of information to the readers, these T cell epitope data recognized in humans /transgenic mouse in case of SARS-CoV that are same/similar to lower ranking T cell epitopes in SARS-CoV2 are provided as supplementary table S1.

Promiscuous helper T cell (HTL) epitopes:

All of the ten SARS-CoV2 proteins, predicted or otherwise, were also studied for helper T cell epitope generation using well validated prediction tool, NetMHCIIpan, in addition to an immunogenicity prediction tool, CD4episcore, which predicts epitopes based on both HLA-binding and immunogenicity. Prominent HLA-II alleles studied using NetMHCIIpan were HLA DRB1 alleles, specifically, DRB1*01:01, DRB1*03:01, DRB1*07:01, DRB1*09:01, DRB1*10:01, DRB1*11:01 and DRB1*15:01, because these alleles are found to be frequent across populations ranging from North America, India, Japan, China, Africa and Europe (allelefrequencies.net). The alleles in CD4episcore studied are: HLA-DRB1:03:01, HLA-DRB1:07:01, HLA-DRB1:15:01, HLA-DRB3:01:01, HLA-DRB3:02:02, HLA-DRB4:01:01 and HLA-DRB5:01:01.

Helper T lymphocyte epitopes are typically 15 amino acids residues long. High throughput data for these epitopes was analysed manually to identify common epitopes across alleles and 10 coronaviral proteins.

From NetMHCIIpan studies, a total of 1802 HTL epitopes (same epitope is predicted to be bound to

multiple alleles) selected till rank 2% which are strong binders (or till rank 10%, weak binders in case strong binders were not found) were generated. Among these epitopes, 649 epitopes (15-mer) were found to be immunogenic by CD4episcore across all alleles. Another immunogenicity prediction tool, ITcell, was used to predict immunogenic epitopes across two alleles DRB1*01:01 and DRB1*15:01 as it uses PDB files for TCR and there was no structure for other alleles in PDB. Also, ITcell predicts 12-mer HTL epitopes. Taking ITcell results into account, top scoring common immunogenic epitopes to both these immunogenicity prediction tools were 95 in number and were taken for further analysis. These also included some of the epitopes for other HLA-DRB1 alleles studied. This can be explained on the basis of observations that among all HLA-II molecules, there exists a high degree of repertoire overlap, reflecting multiple binding partners. This is most probably due to backbone interactions rather than anchor residues playing a major role (18). Many of the top-scoring immunogenic epitopes were common among the two immunogenicity prediction tools, and top 50 high scoring candidate epitopes are tabulated in Table 3. A complete list of these and other epitope candidates are provided in Supplementary Table S2. This list also provides immunogenic HTL epitopes in RBM region (437-508), the ACE-2 binding motif of RBD of Surface protein, which has been shown to elicit neutralizing antibodies (14). The whole dataset of HLA-I and HLA-II epitopes across these mentioned as well as other alleles is available as supplementary information (Supplementary Tables S4, S5 and S6).

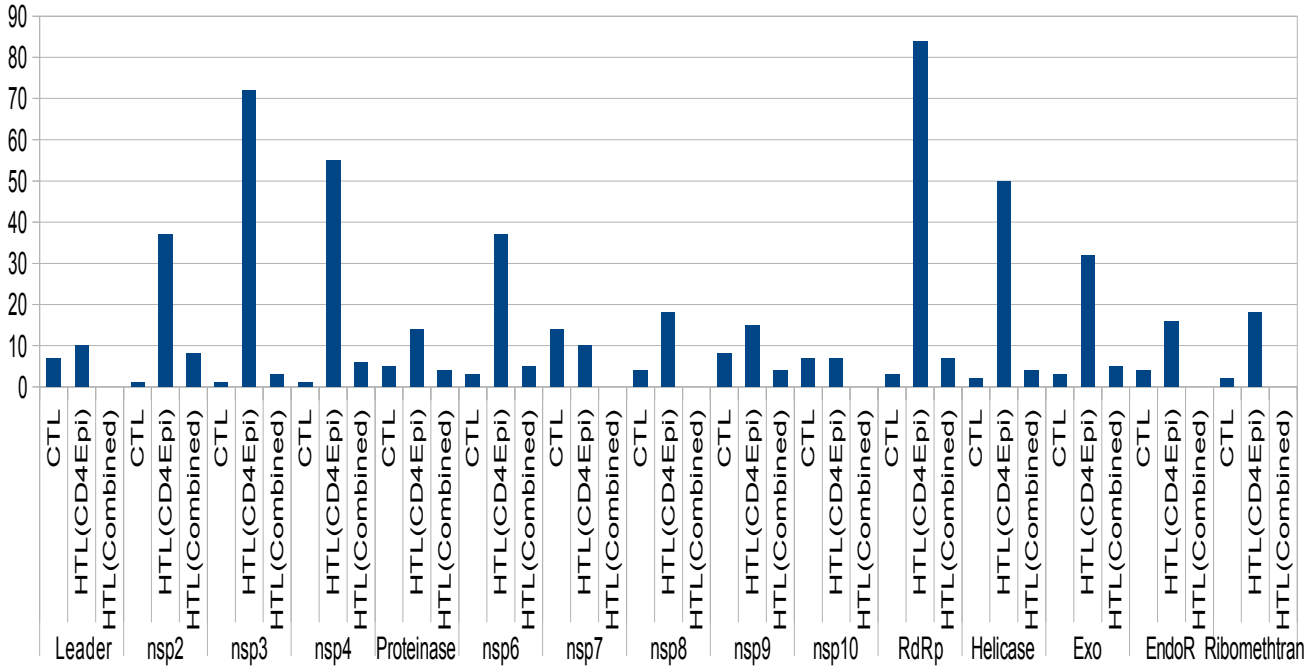
Table 3: Top 50 immunogenic sequences from CD4episcore and ITcell tools, Red Colored fonts: common to IT cell immunogenicity epitopes sorted by DRB1*0101 score, Blue highlights: common to ITcell immunogenicity epitopes sorted by DRB1*1501 score, Yellow highlights: Immunogenic candidates from CD4episcore and common to ITcell and different from Grifoni et al Cell Host and Microbe, 2020 paper with patent; also those in blue highlights that are different from Grifoni etal., 2020 paper have been put into text.

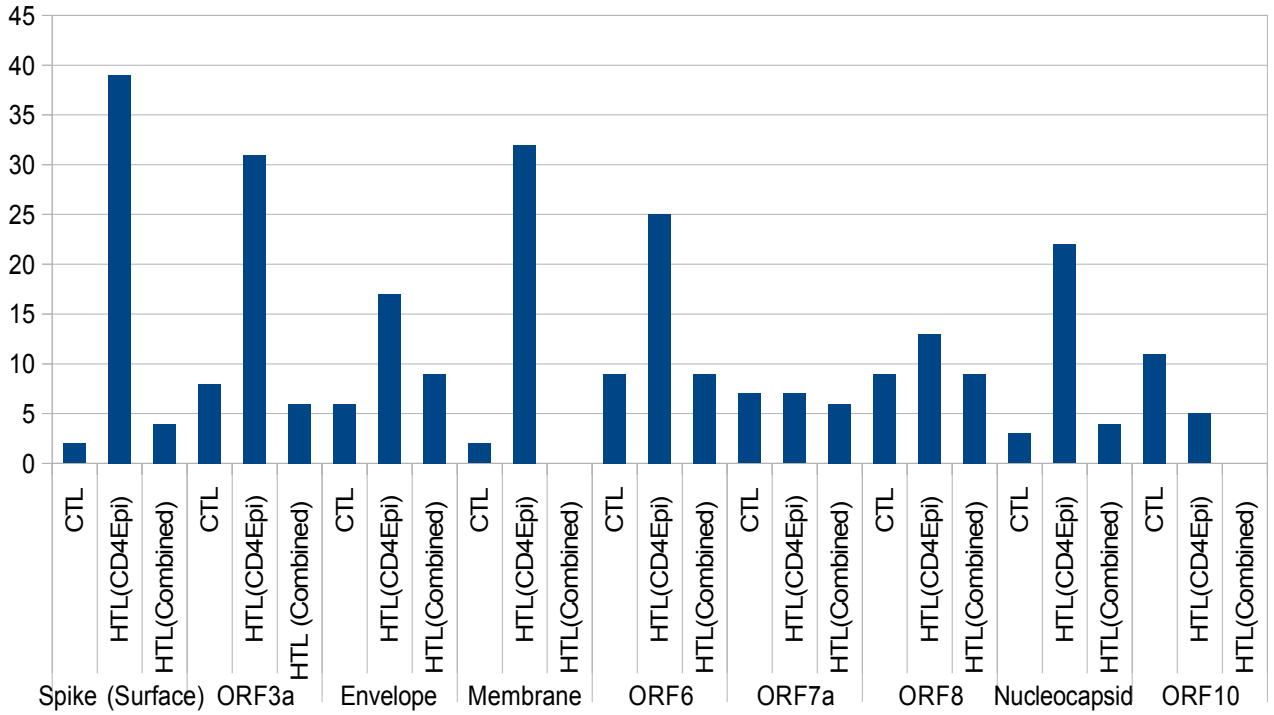
Protein	Protein Numbe	Protein Description	Peptide	Peptide start	Peptide end	Combined Score
Membrane	58	seq47, seq58	SYFIASFRLFARTRS	94	108	22.026
ORF6	28	seq28, seq44	ILLIIMRTFKVSIWN-Differenl	14	28	22.64144
nsp4	106	seq106	FYWFFSNYLKRRVVF	390	404	23.04084
ORF6	22	seq3, seq22, seq25, seq42	EILLIIMRTFKVSIW-different	13	27	23.3002
nsp4	113	seq113	KHFYWFFSNYLKRRV	388	402	23.52856
Membrane	46	seq46, seq59	YFIASFRLFARTRSM	95	109	23.85188
nsp4	109	seq109, seq120	HFYWFFSNYLKRRVV	389	403	24.61324
nsp2	103	seq90, seq103	QTFFKLVNKFLALCA	496	510	24.64816
nsp6	64	seq64	AMMFVKHKHAFLCLF	56	70	24.94928
nsp4	115	seq115	TKHFYWFFSNYLKRR	387	401	24.98624
ORF6	27	seq2, seq27, seq43	AEILLIIMRTFKVSI-different	15	29	25.0036
nsp2	91	seq91, seq111	VQTFFKLVNKFLALC	495	509	25.1072
Membrane	43	seq43, seq64	IASFRLFARTRSMWS	97	111	25.22452
Membrane	42	seq19, seq33, seq39, seq4	ASFRLFARTRSMWSF	98	112	25.24444
nsp6	60	seq60, seq80	AFAMMFVKHKHAFLC	54	68	25.25948
nsp6	81	seq62, seq81	FAMMFVKHKHAFLCL	55	69	25.52472
nsp6	75	seq58, seq75	SAFAMMFVKHKHAFL	53	67	26.23016
nsp2	105	seq89, seq105	SVQTFFKLVNKFLAL	494	508	26.38856
RdRp	5	seq5, seq71, seq100, seq1	MPNMLRIMASLVLAR-differ	626	640	26.47384
Membrane	60	seq45, seq60	FIASFRLFARTRSMW	96	110	26.53024
nsp2	108	seq95, seq108	TFFKLVNKFLALCAD	497	511	26.67152
RdRp	113	seq9, seq75, seq113, seq1	AMPNMLRIMASLVLA-differ	625	639	27.38192
RdRp	53	seq53, seq135	LRIMASLVLARKHTT-differe	630	644	27.69396
RdRp	20	seq20, seq180	RAMPNMLRIMASLVL	624	638	28.1144
nsp4	125	seq125	STKHFYWFFSNYLKR	386	15	28.29724
RdRp	152	seq1, seq48, seq70, seq91,	PNMLRIMASLVLARK-differ	627	641	28.29936
Ribose methyl	6	seq6, seq48	GRLIIRENNRVVISS	278	292	28.37836
RdRp	120	seq11, seq49, seq120, seq	MLRIMASLVLARKHT	629	643	28.38828
RdRp	74	seq6, seq45, seq74, seq10	NMLRIMASLVLARKH-differ	628	642	28.83568
ORF6	45	seq4, seq20, seq29, seq45	LLIIMRTFKVSIWNL-differen	15	29	28.88076
ORF6	48	seq17, seq30, seq48	LIIMRTFKVSIWNLD	16	30	29.11184
RdRp	99	seq19, seq83, seq99, seq1	QMNLKYAISAKNRAR	541	555	29.1294
nsp8	28	seq28	VVLKKLKKSLNVAKS	33	47	29.31276
Ribose methyl	7	seq7, seq47	KGRLIIRENNRVVIS	277	291	29.31836
nsp8	27	seq27	EVVLKKLKKSLNVAK	32	46	29.46256
Ribose methyl	8	seq8	RLIIRENNRVVISSD	279	293	29.7396
nsp2	94	seq94	ESVQTFFKLVNKFLA	493	507	29.9796
RdRp	107	seq86, seq107, seq139	TQMNLKYAISAKNRA	540	554	30.01152
Membrane	68	seq21, seq34, seq38, seq4	SFRLFARTRSMWSFN	99	113	30.13772
Nucleocapsid	49	seq49	DQIGYYRRATRRIRG	82	96	30.45712
RdRp	87	seq24, seq87, seq108, seq	MNLKYAISAKNRART	542	556	30.4728
nsp6	22	seq22, seq92	VLLILMTARTVYDDG-differe	121	135	30.78116
Exonuclease	56	seq15, seq32, seq56	AYNMMISAGFSLWVY	497	511	30.94884
Nucleocapsid	48	seq48	QIGYYRRATRRIRGG	83	97	30.9554
Exonuclease	53	seq11, seq30, seq53	DAYNMMISAGFSLWV	496	510	31.22164
nsp6	91	seq20, seq84, seq91	VVLLILMTARTVYDD-differe	120	134	31.25796
Helicase	84	seq84	CFKMFYKGVITHDVS	471	485	31.35264
Exonuclease	71	seq3, seq31, seq39, seq60,	DMTYRRLISMMGFKM-diffe	48	62	31.65644
Ribose methyl	49	seq9, seq49	SKGRLIIRENNRVVI	276	290	31.83036
Nucleocapsid	50	seq50	IGYYRRATRRIRGGD	84	98	31.83548

Bar diagram for CTL and HTL immunogenic epitope distribution across proteins (Fig. 1) shows a general trend with the number of epitopes not correlated with the size of proteins. The smallest predicted protein, ORF10, is found to provide more number of CTL epitopes in the context of this study than the larger spike protein. Some previous studies have also found this to be true, wherein capsid and matrix proteins in studied viruses were found to "pack significantly more epitopes than those expected by their size" (20) . Some proteins such as ORF6, ORF8, ORF10, Envelope and

Membrane do not have immunogenic HTL epitopes that harbour nonameric CTL epitopes binding to either HLA-DRB1*0101 and HLA-DRB1*1501, and in some cases to none of the two alleles. Also, leader, nsp7, nsp10 and EndoRNAse proteins of ORF1ab did not provide common epitopes between the two immunogenicity prediction tools. The highest number of immunogenic HTL epitopes as predicted by CD4episcore was provided by RdRp, followed by nsp3, nsp4, helicase and spike (surface) protein.

Fig. 1: Bar diagram for CTL and HTL immunogenic epitope distribution across proteins. HTL (CD4epi) depicts immunogenic epitopes from CD4episcore tool. HTL (Combined) depicts epitopes common to CD4episcore and ITcell tools. First panel: All ORF1ab proteins.





Venn diagram showed a common list of many epitopes from a single protein across alleles (Supplementary fig. S2). A distinct pattern is to be noted, analysis of HTL epitopes belonging to HLA-DRB1*03:01, HLA-DRB1*11:01 and HLA-DRB1*15:01 showed the lowest number of common epitopes or none at all across most of the proteins, and can be considered outlier epitopes. Envelope protein was unique in the sense that it did not provide either strong or weak binders to HLA-DRB1*03:01 allele. ORF10 was also unique in providing only weak HTL binders to all of the alleles studied. Venn diagram of all these cytotoxic and helper T cell epitopes taken together showed no common epitopes at all across proteins, but within a protein set, common epitopes can be seen. This shows that every protein of SARS-CoV2 may present antigenic epitopes to the immune system, resulting in a high number of targets. This further lends credence to the theory that multiple T cell epitopes may elicit an immune response in each case, some eliciting strong and some providing weaker responses and therefore, there may be high degree of T cell immunopathology at the infection site. Stronger T cell immune response may cause even the normal, uninfected cells to be attacked while weaker helper T cell immune response, in some protein targets, may cause weak neutralizing antibody

responses as well as weak CTL response at varying times during infection. Very recently, one study has pointed to this immune dysregulation (21) in Covid19 patients with IL6-mediated low HLA-DR expression with sustained cytokine production. Another correspondence paper also pointed to a cytokine storm in context (22). Antibody-mediated enhancement of immune response is also not ruled out and can be seen from the fact that all the epitopes present in the list of dominant B cell epitopes (Table 4 in ref 23) belonging to surface, membrane and nucleocapsid protein, are unique, and there may be a higher non-neutralizing antibody level in Covid19 patients, like in the case of dengue viruses (23).

While this study was at writing stage, two studies on T cell epitope generation using all proteins (24, 25) were published. This present study is different from Grifoni et al. 2020 (24) study in that two prediction tools with very different algorithms, one using neural network and another using postion-specific weight matrices were employed to generate a list of common epitopes, thereby increasing prediction accuracy. Also, Grifoni et al, 2020 focussed mostly on previous SARS coronavirus epitope similarity for predicting epitopes, while this paper identified several novel epitopes across all ten proteins using two different prediction algorithms in each case. Further, this epitope list comprises of common top-scoring epitopes with a higher accuracy and is restricted to highly frequent HLA alleles across population. Also, in view of the several mutations in SARS-CoV2 genome distinct from SARS-CoV, these epitopes not found from SARS-CoV similarity may be potentially more immunogenic. Most of the novel HTL and CTL epitopes were distinct from the epitopes predicted by Grifoni *et al.*, 2020, and were found among top 100 immunogenic candidates predicted by CD4episcore as well as those in common to ITcell predictions (Supplementary table S2). There was no supplementary material on the website or sequence information of the epitopes in the study from Nguyen et al., 2020 (25). Further, their work did not take into account TAP transporter binding predictions as well as HLA-II binding studies, while this study used all three stages of MHC processing and presentation pathway: proteasomal cleavage, TAP transporter binding and MHC class I and II-

binding as well as immunogenicity studies into account for predictions.

Clustering analysis:

All 1924 CTL and HTL top-most epitopes (122 CTL epitopes and 1802 HTL epitopes) across the proteins studied, of which 1096 were non-redundant, unique epitopes, were then clustered using IEDB epitope cluster analysis tool (26) to make further biologically meaningful decisions. Results analyzed suggested that many epitopes were clustered around one consensus sequence (Supplementary table S3). The total number of clusters (including subclusters) was 244, and 66 epitopes were singletons not present in a cluster.

The larger clusters harbouring consensus sequences were:

VDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKN (23 members),

KLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLSM and

TQHQPYVVDDPCPIHFYSKWYIRVGARKSAPLIEL (20 members each). These clusters across proteins and alleles may be considered immunodominant epitopes and tested first among the ranked list of epitopes.

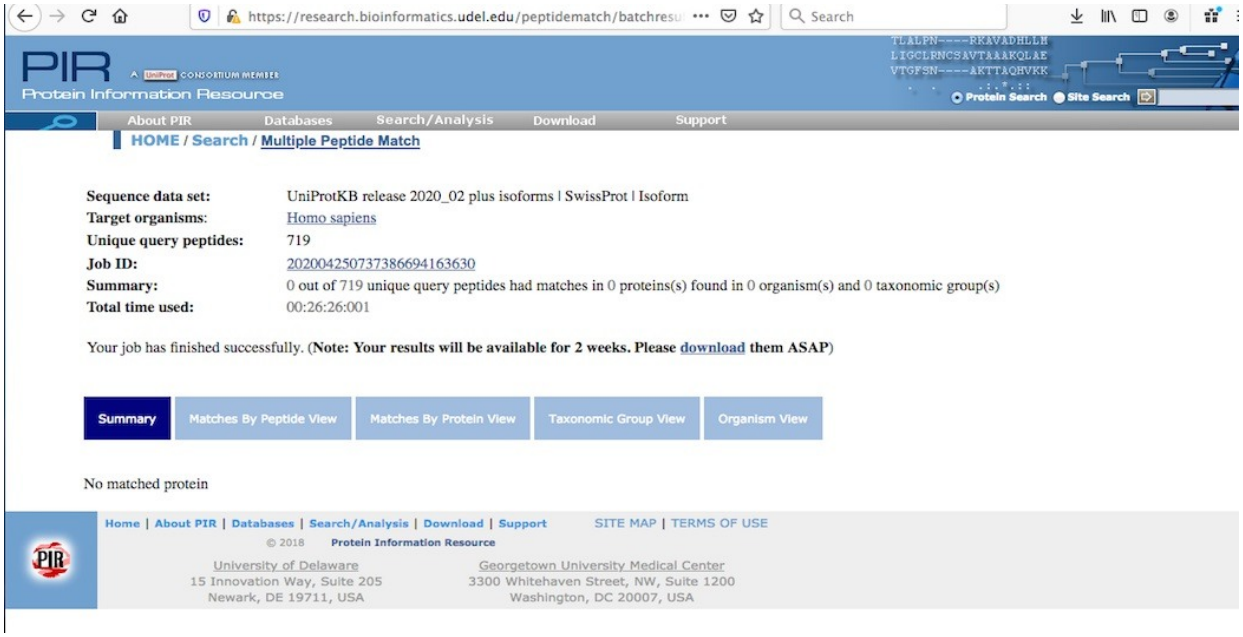
Among immunogenic 122 CTL epitopes from IEDB and 666 HTL epitopes from CD4episcore, again HLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLII topped the list. Further, among the same immunogenic 122 CTL and 95 HTL epitopes common to two prediction algorithms, CD4episcore and ITcell, VTIAEILLIIMRTFKVSIWNLDYIINL belonging to ORF6 again topped. Moreover, PIHFYSKWYIRVGARKSAPLIEL belonging to ORF8 and MGYINVFAFPFTIYSLL belonging to ORF10 were also among the top 3 clustered sequences. It is of interest to note that sequences in the consensus sequence MGYINVFAFPFTIYSLL belonging to ORF10 are weak binders to all the HLA-DRB1 alleles studied while the same sequences are strong binders to all HLA-I supertypes studied.

Table S3: Consensus and singleton sequences generated using IEDB Clustering tool

Crossreactivity studies:

Crossreactivity analyses against human proteome based on UniProt data (Fig. 2) showed that all the immunogenic CTL and HTL epitopes (all HTL epitopes taken from CD4episcore list, removing redundant HTL epitopes; total 719 CTL + HTL epitopes) obtained were not present in human proteome and hence, no crossreactivity to normal human cells may occur.

Fig. 2: Multiple Peptide Match of 719 predicted SARS-CoV2 coronaviral epitopes aganist *Homo sapiens* proteome from UniProt.



B-cell Epitopes:

The widespread presence of novel, unique T cell epitopes in the SARS-Cov2 proteome, is also the main reason that in this paper, B cell epitopes were not studied. Including B cell epitopes in the vaccination strategy with T cell epitopes may not be a good strategy, and may even be counter-productive. Even though neutralizing antibody levels are found to be low in Covid19 patients (27, 28), it is expected that CD4+ T cell expansion responses may increase the neutralizing antibody levels (29) and hence

quantifying CD4+T cell responses using IFN-gamma ELISPOT assays will be useful. This is so done in order to minimize the possible immune system backfiring (21, 22) due to the presence of too many overlapping as well as non-overlapping epitopes in multi-subunit vaccines. It is suggested that helper T cell epitopes be chosen so as to elicit an immune response robust enough to prime and maintain antibody responses, as well as keep the immunopathology under check. In the proven scenario of immune system backfiring , it may be one possible mechanism by which SARS-CoV2 may be acting at its deadliest nature. It is indeed, a dangerous pathogen to control, although for effective immunotherapy at a global scale, efforts should already be underway using these ranked list of epitopes. Almost all of its proteins may pose as foreign agents to the human immune system, with each protein contributing several unique, different immunogenic epitopes. This horde of foreign proteins brings down an avalanche of immune system molecules to the infection site, in order to fight the virus. But instead of immune protection, this may lead to immune enhancement or allergic inflammation at the infection site. These analyses show that coronavirus genome has evolved to be a unique genome. Even as this study is important in pointing out the possible mechanisms in contagious nature of SARS-CoV2 , more evidence is required in the form of experiments.

While many of the proteins studied are found to be expressed and also their functions known by virtue of homology with SARS-CoV, many of the novel ORFs including ORF8 and ORF10 need to be experimentally tested for their expression and functional validation. Experimental MHC-peptide binding and T cell assays are now required for *in vitro* testing for further refinement and development as potent immunogens to be incorporated as components of subunit vaccines.

Conclusions:

Utilizing all ten SARS-CoV2 proteins, predicted or otherwise, a ranked list of CTL and HTL epitopes with high HLA binding affinity, high TAP transport efficiency and high C-terminal proteasomal cleavage ranking has been generated. Utilizing alleles predominant in whole world population, two different prediction algorithms were implemented in identification of common epitopes

for consensus. Immunogenicity scores for these epitopes have also been predicted in order to further narrow down the list to key few epitopes that can be experimentally tested. Peptide matching with human proteome showed no indication of possible crossreactivity. These epitopes are provided to the scientific community for further *in vitro* and *in vivo* assays and saving their time and costs involved in our urgent bid to tackle SARS-CoV2 infections and ensuing death. This work provides esential information for developing prophylactic and therapeutic interventions and for understanding human immune system responses to this virus.

Materials and Methods:

Genome sequence:

The genome sequence of novel coronavirus was retrieved from GenBank accession number MT106054.1/RefSeq sequence number NC_045512.2 and the corresponding proteins were retrieved. RefSeq sequences of all of the proteins present in this genomic sequence, ORF10 protein (YP_009725255.1), nucleocapsid phosphoprotein (YP_009724397.2), ORF8 protein (GenBank: QID21074.1, no RefSeq sequence identified for ORF8), ORF7a protein (YP_009724395.1), ORF6 protein (YP_009724394.1), membrane glycoprotein (YP_009724393.1), envelope protein (YP_009724392.1), ORF3a protein (YP_009724391.1), surface glycoprotein (YP_009724390.1), ORF1ab (YP_009724389.1) were analysed in order to cover the entire genome of SARS-CoV2 in view of absence of data on its virulent proteins. Within ORF1ab (full protein accession number: YP_009724389.1), the accession number of the following proteins taken were as follows: leader protein-YP_009725297.1, nsp2 -YP_009725298.1 , nsp3 -YP_009725299.1, nsp4- YP_009725300.1, 3C-like proteinase -YP_009725301.1, nsp6 -YP_009725302.1, nsp7 -YP_009725303.1, nsp8 -YP_009725304.1, nsp9 -YP_009725305.1, nsp10 -YP_009725306.1, RNA-dependent RNA polymerase -YP_009725307.1, helicase -YP_009725308.1, 3'-to-5' exonuclease -YP_009725309.1,

endoRNAse -YP_009725310.1 and 2'-O-ribose methyltransferase -YP_009725311.1. Fasta sequences of all of these proteins were taken as inputs in several T cell epitope prediction and analysis tools.

Cytotoxic T cell epitopes prediction:

NetCTLpan version 1.1 (http://www.cbs.dtu.dk/services/NetCTLpan/, 30) and PickPocket version 1.1 (http://www.cbs.dtu.dk/services/PickPocket/, 31) were used. All the parameters used were default parameters. Nonameric peptide epitopes were selected. Epitopes from NetCTLpan were ranked according to the combined score using all three different methods, and epitopes from PickPocket algorithm were sorted by affinity (IC_{50} values in nM). In order to increase prediction accuracy, high scoring epitopes common to both these algorithms (among top 10 in PickPocket and same epitopes among high scoring ones in NetCTLpan) were fished out. 12 HLA supertypes (HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*26:01, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*39:01, HLA-B*40:01, HLA-B*58:01, HLA-B*15:01) as present in both algorithms were used (2). For ORF1ab proteins, promiscuous epitopes were selected among top 30 candidates, as not many common epitopes could be found from NetCTLpan and PickPocket.

Helper T cell epitope prediction:

NetMHCIIpan version 3.2 (http://www.cbs.dtu.dk/services/NetMHCIIpan/, 32) was used to predict helper T cell epitopes across several HLA-DRB1 alleles, specifically, DRB1*01:01, DRB1*03:01, DRB1*07:01, DRB1*09:01, DRB1*10:01, DRB1*11:01 and DRB1*15:01. It works on the basis of quantitative MHC-peptide binding affinity data obtained from the Immune Epitope Database. A consensus list of 15 amino acids long ranked epitopes was generated. For generating top ranked

epitopes, these were sorted using descending order of percent rank. Percent rank is normalized prediction score, comparing to prediction of a set of random peptides (32). The epitopes with %rank <2% and <10% were considered strong and weak binders, respectively.

Immunogenicity prediction:

Immunogenicity is a characteristic property of peptide epitopes that can elicit an immune response. High binding affinity to HLA alleles is not a sufficient criterion for high immunogenicity. Therefore, all the epitopes that were generated as a consensus were checked for their immunogenicity. Immune Epitope database (IEDB) immunogenicity tool (http://tools.iedb.org/immunogenicity/, 33) was used to generate a list of immunogenic CTL eptopes. Immunogenicity of a peptide-MHC complex is predicted based on the physicochemical properties of amino acids and their positions in the predicted peptide. Specifically, amino acids with large and aromatic side chains and positions 4-6 are more important to the immunogenicity of the peptide being presented. Ranking was done after sorting from higher to lower immunogenicity score (33). For helper T cell epitopes immunogenicity prediction, CD4episcore (34) and ITcell (35) were used. CD4episcore was developed using neural networks and combines HLA binding and immunogenicity prediction and outputs a list of immunogenic peptides using a combined score. The authors combined immunogenicity and HLA binding scores, using the median percentile rank score (HLA_score) of the 7-allele method (ranging from 0 to 100) and combined it with their neural network-based immunogenicity score. This combined score is calculated as follows:

Combined score: (alpha * Imm score) + ((1-alpha) * HLA_score), where alpha is optimized to 0.4.

The 7 alleles used are: "HLA-DRB1:03:01","HLA-DRB1:07:01","HLA-DRB1:15:01","HLA-DRB3:01:01","HLA-DRB3:02:02","HLA-DRB4:01:01","HLA-DRB5:01:01". The whole HTL epitope

sequence list belonging to each protein was given as an input and IEDB-recommended combined method was selected for scoring. Lower combined scores imply higher immunogenicity according to the authors developing this prediction tool. The immunogenic vs non-immunogenic epitopes cutoff was a combined score of 50 as per CD4episcore paper.

ITcell works on the basis of three stages of MHC-II processing and presentation pathway. These three stages are, in the authors' (35) own words: "....antigen cleavage, MHCII presentation, and TCR recognition. First, antigen cleavage sites are predicted based on the cleavage profiles of cathepsins S, B, and H. Second, for each 12-mer peptide in the antigen sequence we predict whether it will bind to a given MHCII, based on the scores of modeled peptide-MHCII complexes. Third, we predict whether or not any of the top scoring peptide-MHCII complexes can bind to a given TCR, based on the scores of modeled ternary peptide-MHCII-TCR complexes and the distribution of predicted cleavage sites". The scores are given as normalized Z-scores with negative scores implying higher immunogenicity. The epitope sequences as well as PDB files for TCR molecules corresponding to their cognate MHC alleles were given as an input. The PDB ID for files for HLA-DRB1*01:01 and HLA-DRB1*15:01 alleles are 1FYT.pdb and 1YMM.pdb, respectively. PDB files for all other alleles were not available.

Clustering

As globally conserved epitopes are relevant at this time to contain and treat coronavirus infection, clustering approach was used to find patterns among disparate datasets. In order to group epitopes into several clusters, IEDB epitope cluster analysis tool (26) was applied. All the topmost CTL and HTL epitopes across proteins targets were used as inputs with minimum sequence identity threshold as 70%. Cluster-break algorithm was applied for clear representative sequence.

Cross-reactivity analysis:

All the immunogenic CTL and HTL epitopes obtained were used to search against human proteome

data from UniProt database (2020_02 release, 181,292,975 sequences as of date 06-05-2020) for any matches to human proteome, thus avoiding cross-reactivity. For this, Multiple Peptide Match tool (https://research.bioinformatics.udel.edu/peptidematch/batchpeptidematch.jsp) of Protein Information Resource was used.

Multiple Sequence Alignment:

MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) was used to generate multiple sequence alignments of all SARS-CoV2 proteins with corresponding proteins in other HCoV and MERS species. The species chosen and their GenBank accession IDs were: Alpha-CoV: HCoV-NL63 (NC_005831.2), HCoV-229E (NC_002645.1); Beta-CoV: HCoV-OC43 (NC_006213.1), HCoV-HKU1 (NC_006577.2), MERS CoV (NC_019843.3) and SARS-CoV2 (SARS-CoV2, accession IDs same as above). Spike protein sequence for SARS-CoV was taken from UniProt (P59594). In view of different/unclear annotations, it was difficult to get corresponding protein sequences from SARS-CoV (RefSeq accession ID NC_004718.3). There are no human CoVs in gamma/delta CoV categories. In addition, bat coronavirus RaTG13 sequences (MN996532.1) were also used.

Acknowledgments:

This author acknowledges the tireless help of researchers working towards SARS-CoV2 control and submitting data to GenBank without which these sequence analyses using Immunoinformatics would not have been possible.

Conflict of interest: This author declares that there is no conflict of interest.

References:

References:

1. Smith, Micholas; Smith, Jeremy C. (2020). Repurposing Therapeutics for COVID-

19:Supercomputer-Based Docking to the SARS-CoV-2 Viral Spike Protein and Viral Spike Protein-Human ACE2Interface. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.11871402.v4

2. Seema Mishra and Subrata Sinha (2009). 'Immunoinformatics and modeling perspective of T cell epitope-based cancer immunotherapy: a holistic picture' J Biomol Struct Dyn. 27(3), pp.293-306.
3. Seema Mishra and Subrata Sinha (2006). 'Prediction and molecular modeling of T cell epitopes derived from placental alkaline phosphatase for use in cancer immunotherapy'. J Biomol Struct Dyn. 24(2), pp.109-121.
4. WHO: https://www.who.int/blueprint/priority-diseases/key-action/Novel-Coronavirus_Landscape_nCoV-4april2020.pdf?ua=1
5. Wang, Y-D Fion Sin W-Y, Xu, G-B et al., (2004) T-Cell Epitopes in Severe Acute Respiratory Syndrome (SARS) Coronavirus Spike Protein Elicit a Specific T-Cell Immune Response in Patients Who Recover from SARS. Journal of Virology, 78: 5612–5618
6. Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M, et al. (2015) Epitope-Based Vaccine Target Screening against Highly Pathogenic MERS-CoV: An In Silico Approach Applied to Emerging Infectious Diseases. PLoS ONE 10(12): e0144475. doi:10.1371/journal.pone.0144475
7. Anthony R. Fehr and Stanley Perlman (2015). Coronaviruses: An Overview of Their Replication and Pathogenesis, Methods Mol Biol. 1282: 1–23.
8. Wrapp D, Wang N, Corbett, KS, Goldsmith, JA et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation Science 367, pp. 1260-1263 DOI: 10.1126/science.abb2507
9. Narayanan, K., Huang, C., & Makino, S. (2008). SARS coronavirus accessory proteins. Virus Research, 133(1), 113–121. https://doi.org/10.1016/j.virusres.2007.10.009

10. Kim et al., 2020, The Architecture of SARS-CoV-2 Transcriptome Cell 181, 1–8 https://doi.org/10.1016/j.cell.2020.04.011
11. Oh HL, Chia A, Chang CX, Leong HN, Ling KL, Grotenbreg GMet al. Engineering T cells specific for a dominant severe acute respiratory syndrome coronavirus CD8 T cell epitope. J Virol 85: 10464–10471. (2011).
12. Vartak, A and Sucheck, S. J. Recent Advances in Subunit Vaccine Carriers. Vaccines 4, 12; doi:10.3390/vaccines4020012 (2016).
13. Mishra, S. T Cell Epitope-Based Vaccine Design for Pandemic Novel Coronavirus 2019-nCoV. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.12029523.v2 (2020).
14. Korber B, Fischer WM, Gnanakaran S, et al. (2020) Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. bioRxiv 2020.04.29.069054; doi: https://doi.org/10.1101/2020.04.29.069054
15. Quinlan BD, Mou H, Zhang L, Guo Y, He W, Ojha A, et al., The SARS-CoV-2 receptor-binding domain elicits a potent neutralizing response without antibody-dependent enhancement bioRxiv 2020.04.10.036418; doi: https://doi.org/10.1101/2020.04.10.036418.
16. Wang, C., Li, W., Drabek, D. et al. A human monoclonal antibody blocking SARS-CoV-2 infection. Nat Commun 11, 2251 (2020). https://doi.org/10.1038/s41467-020-16256-y
17. Janice Oh H-L, Gan S K-E, Bertoletti A and T Y-J (2012) Understanding the T cell immune response in SARS coronavirus infection. Emerging Microbes and Infections(2012)1,e23; doi:10.1038/emi.2012.26.
18. Li CK, Wu H, Yan H, Ma S, Wang L, Zhang M et al. T cell responses to whole SARS coronavirus in humans. J Immunol 2008; 181: 5490–5500.
19. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics 2011; 63:325–35.

20. Diez-Rivero CM, Reche PA (2012) CD8 T Cell Epitope Distribution in Viruses Reveals Patterns of Protein Biosynthesis. PLoS ONE 7(8): e43674. https://doi.org/10.1371/journal.pone.0043674

21. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, et al. Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure [published online ahead of print, 2020 Apr 17]. Cell Host Microbe. 2020;S1931-3128(20)30236-5. doi:10.1016/j.chom.2020.04.009

22. Mehta P, McAuley, DF et al., COVID-19: consider cytokine storm syndromes and immunosuppression. The Lancet 395: 10229, P1033-1034, (2020).

23. Dejnirattisai W, Jumnainsong A, Onsirisakul, N et al., Cross-Reacting Antibodies Enhance Dengue Virus Infection in Humans Science 328, 745-748 (2010)

24. Grifoni et al., A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2, Cell Host & Microbe (2020), https://doi.org/10.1016/j.chom.2020.03.002

25. Nguyen, A Julianne K. David, J K, Maden SK et al., Human leukocyte antigen susceptibility map for SARS-CoV-2. J. Virol. doi:10.1128/JVI.00510-20.

26. Sandeep Kumar Dhanda, Kerrie Vaughan, Veronique Schulten, Alba Grifoni, Daniela Weiskopf, John Sidney, Bjoern Peters, Alessandro Sette: Development of a novel clustering tool for linear peptide sequences . Immunology (2018) doi:https://doi.org/10.1111/imm.12984

27. Fan Wu et al. Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered 1 patient cohort and their implications, MedRxiv. https://doi.org/10.1101/2020.03.30.20047365 (2020).

28. Haveri, A. et al. Serological and molecular findings during SARS-CoV-2 infection: the first case study in Finland, January to February 2020 Euro Surveill. 25(11): 2000266 (2020).

29. Nayak JL, Fitzgerald TF, Richards KA, et al. CD4+ T-cell expansion predicts neutralizing

antibody responses to monovalent, inactivated 2009 pandemic influenza A(H1N1) virus subtype H1N1 vaccine. J Infect Dis. 2013 Jan 15;207(2):297-305. doi: 10.1093/infdis/jis684.

30. Stranzl, T., Larsen, M. V., Lundegaard, C., & Nielsen, M. . NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics, 62(6), 357–368 (2010). https://doi.org/10.1007/s00251-010-0441-4

31. Zhang, H., Lund, O., & Nielsen, M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. Bioinformatics (Oxford, England), 25(10), 1293–1299. https://doi.org/10.1093/bioinformatics/btp137 (2009).

32. 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, 65(10), 711–724. https://doi.org/10.1007/s00251-013-0720-y (2013).

33. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Kesmir C, Peters B. Properties of MHC class I presented peptides that enhance immunogenicity. PloS Comp. Biol. 8(1):361. (2013)

34. Dhanda SK, Karosiene E, Edwards L, et al., Predicting HLA CD4 immunogenicity in human populations. Frontiers in Immunology 2018 doi: 10.3389/fimmu.2018.01369

35. Schneidman-Duhovny D, Khuri N, Dong GQ, Winter MB, Shifrut E, Friedman N, et al. (2018) Predicting CD4 T-cell epitopes based on antigen cleavage, MHCII presentation, and TCR recognition. PLoS ONE 13(11): e0206654.