

Cell Host & Microbe, Volume 27

Supplemental Information

A Sequence Homology and Bioinformatic Approach

Can Predict Candidate Targets for Immune

Responses to SARS-CoV-2

Alba Grifoni, John Sidney, Yun Zhang, Richard H. Scheuermann, Bjoern Peters, and Alessandro Sette

Table S2. Predicted B cell epitopes from SARS-CoV-2 spike glycoprotein, membrane glycoprotein and nucleocapsid phosphoprotein using BebiPred 2.0. Related to the SARS-CoV-2 B cell epitope prediction STAR Method section.

Protein	Start	End	Peptide	Length
Spike glycoprotein	10	22	LVSSQCVNLTRT	13
	65	75	FHAIHVSGTNG	11
	109	126	TLDSKTQSLIVNNATNV	18
	146	168	HKNNKSWMESEFRVYSSANNCTF	23
	251	262	PGDSSSGWTAGA	12
	280	288	NENGTITDA	9
	312	320	IYQTSNFRV	9
	356	376	KRISNCVADYSVLYNSASFST	21
	434	444	IAWNSNNLDSK	11
	469	483	STEIYQAGSTPCNGV	15
	526	556	GPKKSTNLVKNKCVNFNFNGLTGTGVLTESN	31
	592	620	FGGVSVITPGTNTSNQVAVLYQDVNCTEV	29
	634	647	RVYSTGSNVFQTRA	14
	652	661	GAEHVNNSYE	10
	673	713	SYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNNSIA	41
	719	739	TISVTTEILPVSMTKTSVDCT	21
	741	755	YICGDSTECNLLLQ	15
	757	769	GSFCTQLNRALTG	13
	771	780	AVEQDKNTQE	10
	867	880	DEMIAQYTSALLAG	14
	915	946	VLYENQKLIANQFNSAIGKIQDLSSTASALG	32
	948	975	LQDVVNQNAQALNTLVKQLSSNFGAISS	28
	1001	1009	LQSLQTYVT	9
1019	1036	RASANLAATKMSECVLGQ	18	
1117	1127	TDNTFVSGNCD	11	
1157	1164	KNHTSPDV	8	
1171	1180	GINASVVNIQ	10	
1188	1197	EVAKNLNESL	10	
1229	1239	MVTIMLCCMTS	11	
Membrane glycoprotein	105	117	RTRSMWSFNPETN	13
	168	195	ITVATSRTLSYYKLGASQRVAGDSGFAA	28
	203	218	NYKLNTDHSSSSDNIA	16
Nucleocapsid phosphoprotein	22	31	DSTGSNQNGE	10
	177	215	RGGSQASSRSSRSRNSRNSTPGSSRGTS ParmAGNGG	39
	232	251	SKMSGKGGQQGQTVTKKSA	20
	405	416	KQLQQSMSSADS	12

Table S4. Prediction of CD4 T cells epitopes listed by SARS-COV-2 proteins. Related to the SARS-CoV-2 T cell epitope prediction STAR Method section.

SARS-CoV-2 CD4 Megapool*		protein length	Percentage of composition	number of epitopes
orf1ab polyprotein	nsp1	180	0.83	2
	nsp2	638	5.81	14
	PLpro	1945	14.11	34
	nsp4	500	9.13	22
	3CL	306	2.49	6
	nsp6	290	7.47	18
	nsp7	83	1.66	4
	nsp8	198	2.07	5
	nsp9	113	0.83	2
	nsp10	138	0.83	2
	RdRpol	933	7.88	19
	Hel	601	5.81	14
	nsp14	527	7.88	19
	nsp15	346	1.66	4
	nsp16	297	3.73	9
Surface glycoprotein	1273	8.30	20	
ORF3a	275	4.15	10	
envelope protein	75	3.32	8	
membrane glycoprotein	222	3.32	8	
ORF6	61	2.49	6	
ORF7a	121	1.66	4	
ORF8	121	1.24	3	
nucleocapsid phosphoprotein	419	2.90	7	
ORF10	38	0.41	1	
Total	9700	100	241	

Table S5. Top 12 HLA class I alleles with a worldwide population frequency $\geq 6\%$. Related to the SARS-CoV-2 T cell epitope prediction STAR Method section.

HLA alleles	Frequency in worldwide population
HLA-A*01:01	16.2
HLA-A*02:01	25.2
HLA-A*03:01	15.4
HLA-A*11:01	12.9
HLA-A*23:01	6.4
HLA-A*24:02	16.8
HLA-B*07:02	13.3
HLA-B*08:01	11.5
HLA-B*35:01	6.5
HLA-B*40:01	10.3
HLA-B*44:02	9.2
HLA-B*44:03	7.6