Cell Reports Medicine, Volume *2*

Supplemental Information

Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases

Alison Tarke, John Sidney, Conner K. Kidd, Jennifer M. Dan, Sydney I. Ramirez, Esther Dawen Yu, Jose Mateus, Ricardo da Silva Antunes, Erin Moore, Paul Rubiro, Nils Methot, Elizabeth Phillips, Simon Mallal, April Frazier, Stephen A. Rawlings, Jason A. Greenbaum, Bjoern Peters, Davey M. Smith, Shane Crotty, Daniela Weiskopf, Alba Grifoni, and Alessandro Sette

Table S1. Characteristics of donor cohorts utilized in the protein screen and utilized to validate megapools. Related to Fig. 1 and Fig. 5-7.

	Protein Screen Cohort	Megapool Validation Cohorts	
	COVID-19 (n = 99)	**Unexposed (n = 25)**	**COVID-19 (n = 31)**
Age (years)	19-91 [median = 41, IQR = 23]	24-82 [Median = 36, IQR = 31]	21-81 [Median = 38, IQR = 29]
Gender			
Male (%)	41% (42/99)	52% (13/25)	35% (11/31)
Female (%)	59% (58/99)	48% (12/25)	65% (20/31)
Sample Collection Date	Mar-Sep 2020	March - May 2020	April - Sept 2020
SARS-CoV-2 PCR	Positive = 100% (78/78)	N/A	Positive 100% (18/18)
	Not tested= 21% (21/99)		Not tested= 42% (13/31)
S RBD IgG	98% (97/99)	0% (0/25)	100% (31/31)
Peak disease Severity[a]			
Mild	91% (90/99)	N/A	58% (18/31)
Moderate	2% (2/99)	N/A	36% (36/31)
Severe	1% (1/99)	N/A	6% (2/31)
Critical	6% (6/99)	N/A	0% (0/31)
Race-Ethnicity			
White- not Hispanic or Latino	76% (76/99)	44% (11/25)	81% (25/31)
Hispanic or Latino	12% (12/99)	8% (2/25)	13% (4/31)
Asian	5% (5/99)	12% (3/25)	3% (1/31)
American Indian/Alaska Native	1% (1/99)	0% (0/25)	0% (0/31)
Native Hawaiian or other Pacific Islander	0% (0/99)	4% (1/25)	0% (0/31)
Black or African American	2%(2/99)	0% (0/25)	0% (0/31)
More than one race	3% (3/99)	0% (0/25)	0% (0/31)
Not reported	0% (0/99)	32% (8/25)	3% (1/31)
Days Post Symptom Onset at Collection[b]	3-184 (108/108) [Median = 67, IQR = 48]	N/A	22-128 (31/31) [Median = 69, IQR = 50]

[a]According to WHO criteria.

[b]Multiple visits for the same donor have been analyzed.

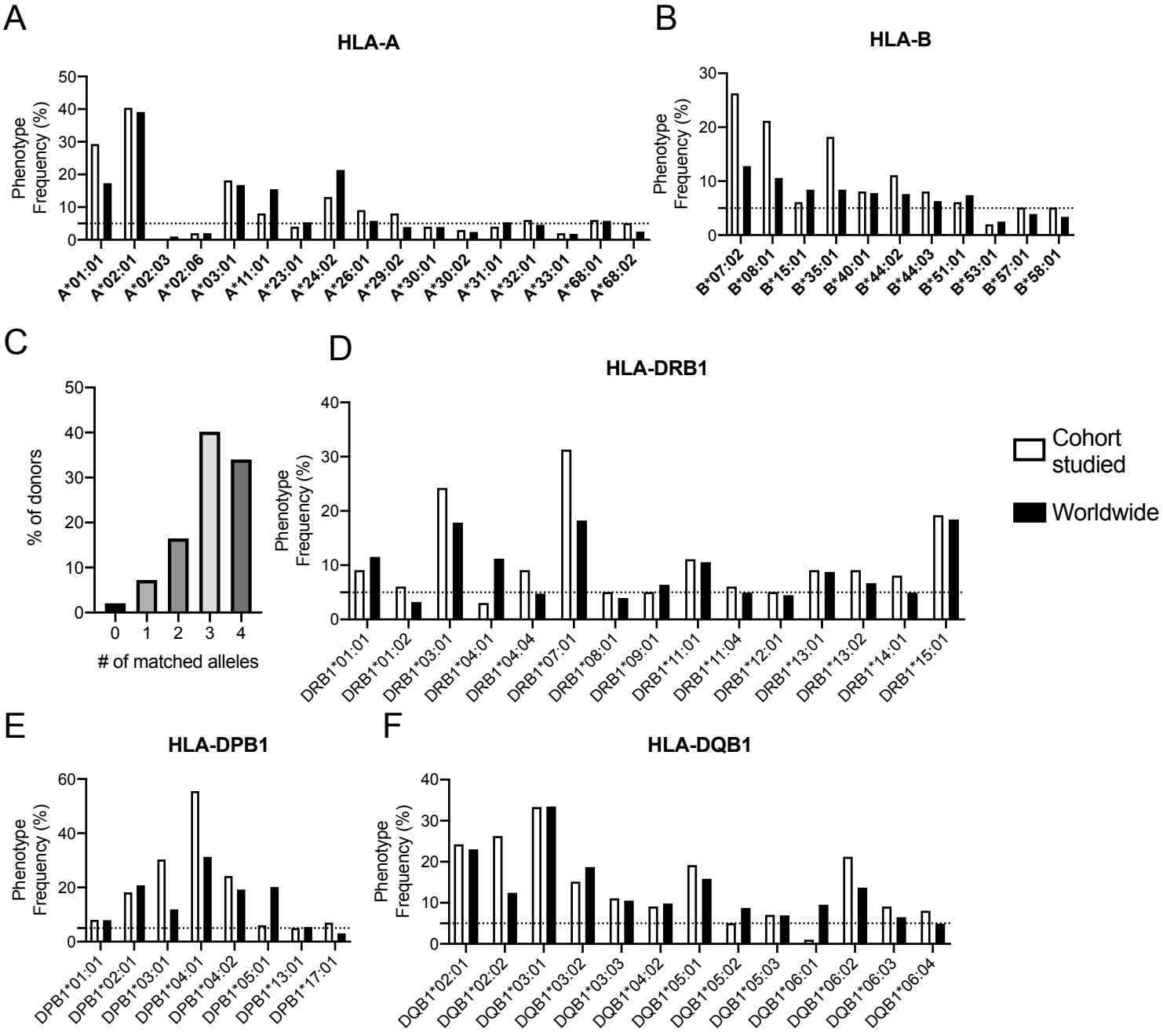

Fig. S1. **Analyses of HLA phenotype frequencies. Related to Fig. 4-5.**
HLA phenotype frequency in the COVID-19 cohort (n = 99) analyzed compared with the worldwide phenotype frequencies available in the IEDB-AR population coverage tool (Bui et al., 2006; Dhanda et al., 2019). HLA class I frequency for A and B loci for the top 28 HLA class I with frequency >5% in the worldwide population are shown in panels (**A**) and (**B**), respectively. Coverage of class I predicted peptides based on the HLA typing of the population (**C**). HLA class II frequency for DRB1, DP and DQ loci for the top HLA class II with frequency >5% in the worldwide population of the studied cohort are shown in panels (**D**), (**E**), and (**F**) respectively.

Table S3. Summary of SARS-CoV-2-specific T cell reactivity as a function of the most immunodominant proteins.[a] Related to Fig. 1.

CD4$^+$ T cells (n=99)								**CD8$^+$ T cells (n=99)**							
proteins	Frequency (%)	Total counts	Average counts	% of responses per protein	Cumulative (%)	# of peptides tested	Cumulative (%)	proteins	Frequency (%)	Total counts	Average counts	% of responses per protein	Cumulative (%)	# of peptides tested	Cumulative (%)
S	**95**	**471873**	**4766**	**26**	**25**	**253**	**13**	**S**	**80**	**440927**	**4454**	**26**	**26**	**253**	**13**
M	**96**	**268883**	**2716**	**17**	**43**	**43**	**15**	**nsp3**	**49**	**174104**	**1759**	**17**	**43**	**388**	**33**
N	**93**	**220425**	**2227**	**15**	**58**	**82**	**20**	**N**	**75**	**180191**	**1820**	**15**	**58**	**82**	**38**
nsp3	**70**	**160567**	**1622**	**8**	**66**	**388**	**40**	**M**	**69**	**149071**	**1506**	**11**	**69**	**43**	**40**
ORF3a	**76**	**94025**	**950**	**5**	**70**	**53**	**43**	**ORF3a**	**47**	**58439**	**590**	**4**	**73**	**53**	**43**
nsp12	**64**	**75157**	**759**	**4**	**74**	**186**	**52**	**nsp4**	**45**	**39945**	**403**	**3**	**76**	**100**	**48**
nsp4	**58**	**68778**	**695**	**3**	**77**	**100**	**57**	**nsp12**	**35**	**26764**	**270**	**3**	**79**	**186**	**57**
nsp13	**58**	**51681**	**522**	**3**	**80**	**120**	**64**	**nsp6**	**34**	**26619**	**269**	**2**	**81**	**58**	**60**
ORF8	**58**	**50217**	**507**	**2**	**83**	**23**	**65**	nsp2	30	22324	225	2	83	127	67
nsp16	34	29175	295	2	85	59	68	nsp1	24	18393	186	2	85	36	69
nsp2	54	48674	492	2	87	127	74	nsp5	31	17136	173	2	87	61	72
ORF7a	48	59270	599	2	89	23	76	nsp16	20	20221	204	2	89	59	75
Nsp6	49	28783	291	2	91	58	79	nsp13	25	16004	162	2	91	120	81
nsp14	44	26059	263	2	93	105	84	ORF8	32	22389	226	1	92	23	83
nsp5	46	29139	294	2	94	61	87	ORF7a	28	19612	198	1	93	23	84
E	33	21161	214	1	95	13	88	nsp14	19	11135	112	1	94	105	89
nsp1	22	20357	206	1	97	36	90	nsp7	18	7480	76	1	95	17	90
nsp8-9-10	37	21419	216	1	98	91	95	nsp8-9-10	18	14067	142	1	97	91	95
nsp15	29	17154	173	1	98	70	98	ORF6	24	12947	131	1	98	11	95
ORF6	27	10632	107	1	99	11	99	nsp15	14	12208	123	1	99	70	99
ORF10	19	7551	76	0	100	6	99	ORF10	20	12178	123	1	99	6	99
nsp7	19	14248	144	0	100	17	100	E	20	9542	96	1	100	13	100

[a]Bold font indicates the top SARS-CoV-2 proteins accounting for a cumulative response >80% for CD4$^+$ and CD8$^+$ T cells.

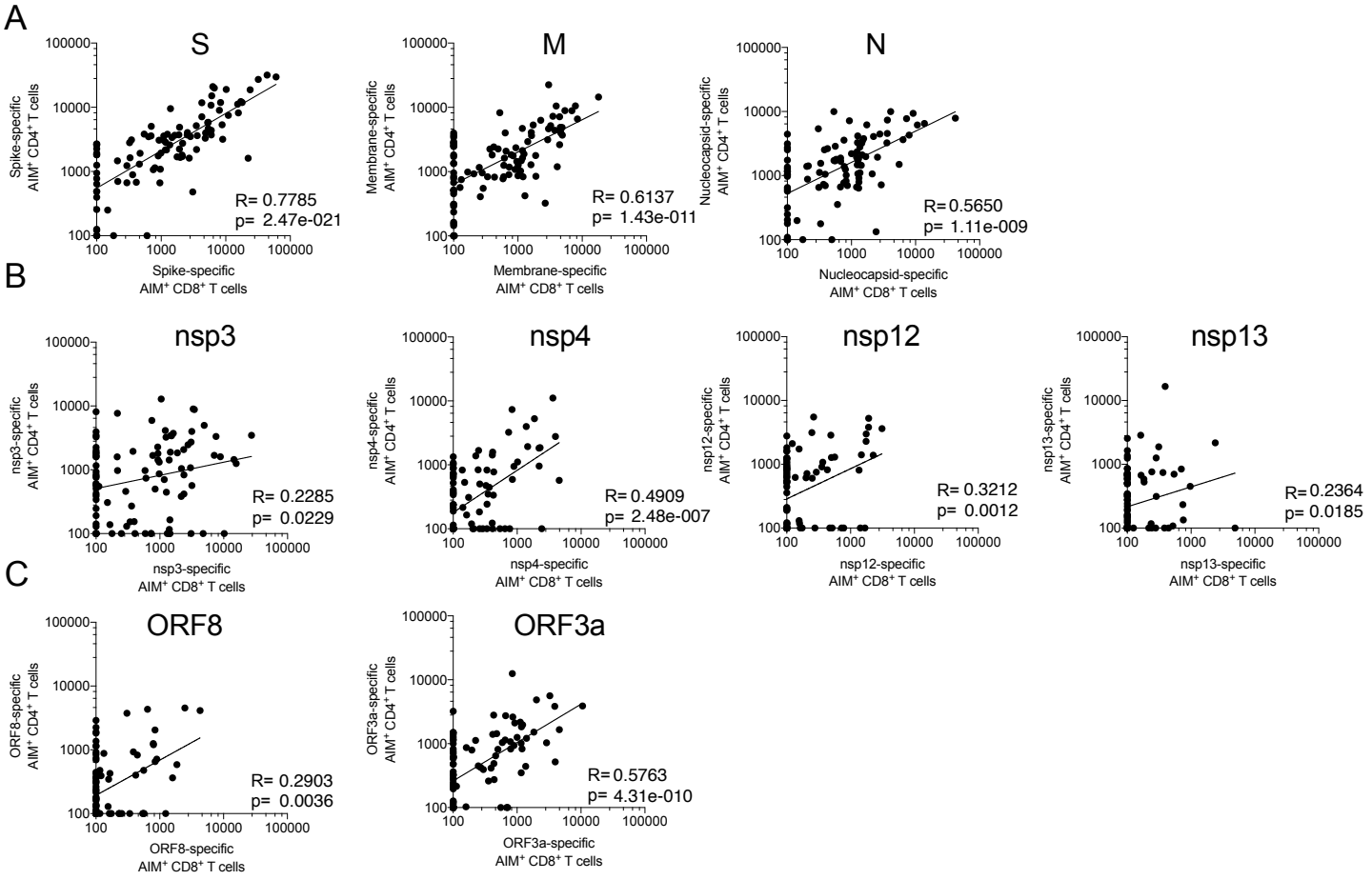

Fig. S2. **Correlations of SARS-CoV-2-specific CD4$^+$ and CD8$^+$ T cell reactivities per protein. Related to Fig. 2.**
CD4$^+$ and CD8$^+$ T cell reactivities are correlated for each of the 9 SARS-CoV-2 antigens that were immunodominant for CD4$^+$ T cells: S, M, and N (**A**); nsp3, nsp4, nsp12, and nsp13 (**B**); and ORF8 and ORF3a (**C**). All analyses were performed using Spearman correlation and the p-values shown were not corrected for multiple hypothesis testing.

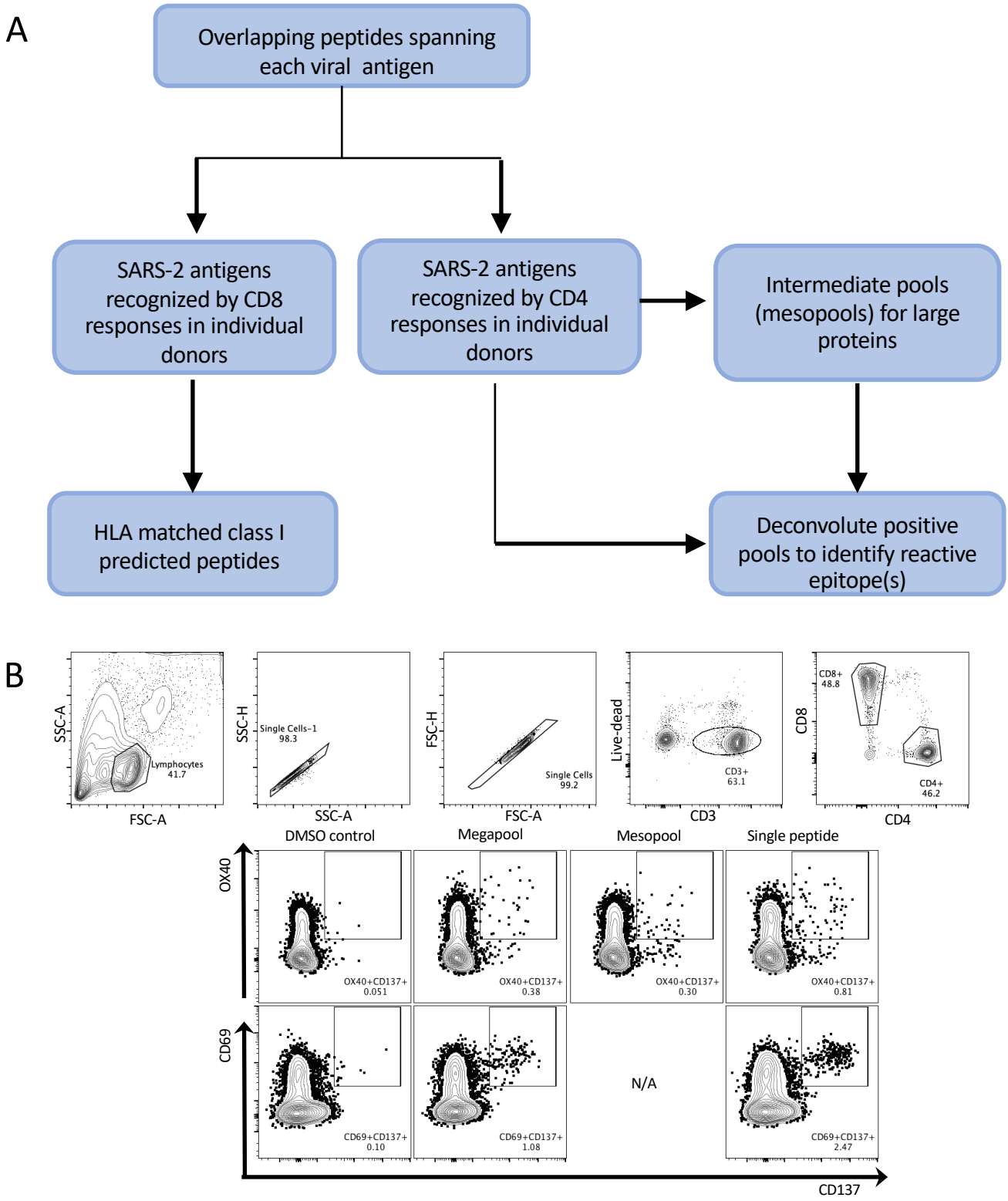

Fig. S3. **Summary of experimental strategy. Related to Fig. 1.**
Scheme of experimental strategy selected for HLA class I and class II epitope identification (**A**). Representative graphs depicting the flow cytometry gating strategy for defining antigen-specific $CD4^+$ and $CD8^+$ T cells by $OX40^+CD137^+$ and $CD69^+CD137^+$ expression, respectively (**B**).

Table S4. Number of predicted epitopes synthetized based on the most frequent 28 HLA class I alleles. Related to Fig. 4.

Allele	nsp1	nsp2	nsp3	nsp4	nsp5	nsp6	nsp7	nsp8	nsp9	nsp10	nsp12	nsp13	nsp14	nsp15	nsp16	S	ORF3a	E	M	ORF6	ORF7a	ORF8	N	ORF10	Total per allele
A*01:01	3	9	36	7	9	0	0	5	21	1	39	11	4	2	3	26	9	0	7	0	2	2	4	0	200
A*02:01	4	12	35	12	7	16	3	5	2	1	17	11	12	10	6	17	10	2	4	3	5	3	2	1	200
A*02:03	8	18	39	16	6	11	3	5	2	2	13	12	11	4	6	25	6	3	1	1	1	5	2	0	200
A*02:06	3	9	42	18	4	12	3	3	2	1	20	10	12	12	7	18	8	2	2	3	2	3	2	2	200
A*03:01	4	14	51	7	6	0	2	7	1	1	23	18	5	9	6	30	5	0	2	1	0	0	8	0	200
A*11:01	1	17	57	9	3	3	1	7	1	1	21	10	5	9	5	25	8	1	5	0	1	0	10	0	200
A*23:01	2	4	45	16	3	14	0	2	2	0	19	6	15	7	6	34	8	0	4	0	1	3	5	4	200
A*24:02	2	4	48	14	3	16	0	2	3	0	20	6	16	8	5	35	6	0	4	0	1	1	2	4	200
A*26:01	3	8	61	11	4	3	2	2	1	2	30	17	8	5	5	23	6	0	2	0	1	1	4	1	200
A*29:02	1	9	43	17	6	8	0	2	1	1	27	17	11	6	3	23	8	2	6	1	1	3	2	2	200
A*30:01	3	12	43	6	3	3	0	4	1	1	24	18	10	9	6	21	4	3	9	0	3	1	15	1	200
A*30:02	2	4	43	13	4	6	0	5	2	0	29	16	10	6	2	29	5	3	9	1	1	3	6	1	200
A*31:01	6	13	20	11	4	5	0	3	2	0	31	13	9	7	6	33	2	3	15	0	3	2	12	0	200
A*32:01	2	11	42	17	1	11	6	6	0	0	16	13	13	7	9	23	8	2	4	2	1	1	4	1	200
A*33:01	9	10	19	16	7	6	0	4	4	0	25	15	13	8	7	25	5	1	14	0	3	3	5	1	200
A*68:01	5	15	46	10	4	8	0	5	0	2	26	11	4	10	6	25	5	2	8	0	1	0	7	0	200
A*68:02	1	10	48	10	2	8	2	4	0	0	23	17	8	4	4	31	10	4	3	2	2	2	4	1	200
B*07:02	8	9	35	9	9	3	0	7	3	3	17	23	6	11	9	21	4	0	4	0	4	1	12	2	200
B*08:01	3	14	26	13	7	6	3	5	4	1	27	15	12	5	9	26	7	1	6	2	3	0	4	1	200
B*15:01	3	11	46	15	4	5	3	6	0	4	26	13	13	7	2	22	0	2	6	0	2	2	8	0	200
B*35:01	4	4	44	12	2	8	2	4	3	3	26	10	11	9	7	29	4	1	4	0	4	0	8	1	200
B*40:01	5	27	53	13	1	2	2	8	1	3	16	10	5	16	6	19	1	1	3	1	3	3	1	0	200
B*44:02	5	25	53	5	1	1	2	8	1	2	20	12	8	7	8	19	8	0	4	1	1	0	9	0	200
B*44:03	5	24	57	5	1	1	2	8	1	3	19	12	7	8	8	16	8	0	4	1	2	0	8	0	200
B*51:01	5	16	35	12	4	11	1	6	1	1	22	15	6	7	7	28	7	0	2	0	1	2	7	4	200
B*53:01	3	8	37	5	5	10	2	4	2	3	18	9	15	9	4	39	5	0	4	0	5	1	10	2	200
B*57:01	3	8	44	9	8	17	2	7	2	0	14	9	13	8	10	17	10	0	8	1	5	0	5	0	200
B*58:01	3	5	42	8	8	13	1	9	2	0	18	8	15	9	12	19	12	0	7	1	2	0	5	1	200
Total per protein	106	330	1190	316	126	207	42	143	65	36	626	357	277	219	174	698	179	33	151	21	61	42	171	30	5600

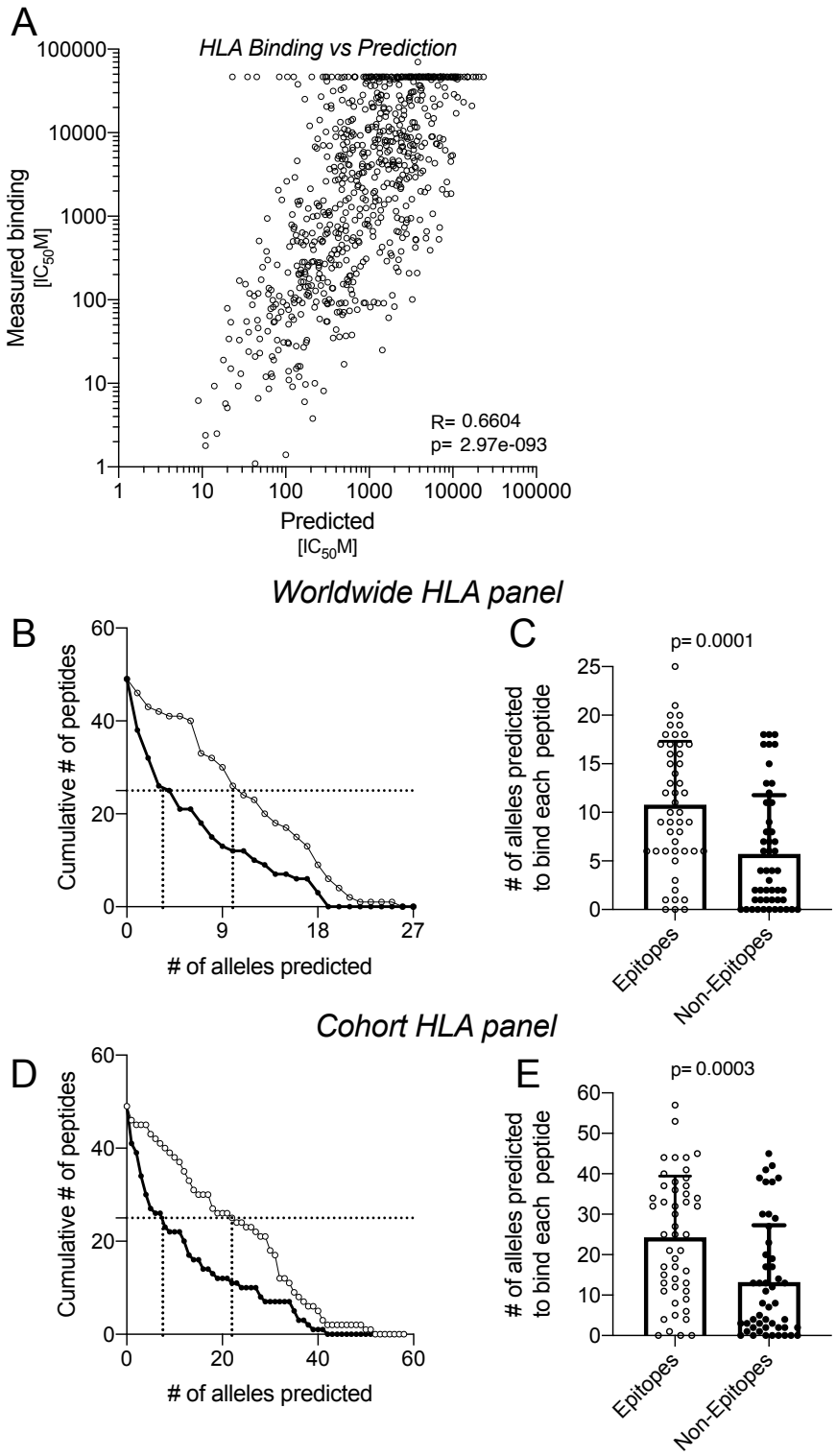

Fig. S4. **SARS-CoV-2 immunodominant epitope HLA class II binding capacity and promiscuity. Related to Fig. 3.**
A comparison of the HLA class II binding capacity of 49 immunodominant epitopes as determined by binding predictions or as measured experimentally (**A**), suggesting feasibility for using binding predictions to assess HLA-restriction. Predicted HLA class II binding promiscuity is shown for the same 49 epitopes (white circles), and also 49 non-epitopes (black circles), considering the 27 HLA class II alleles most frequent worldwide (**B-C**), or the 58 HLA class II alleles specific to the study cohort (**D-E**).The number of HLA class II alleles predicted to bind epitopes (white circles) and non-epitopes (black circles) are based on a prediction cutoff value of IC_{50}≤1000nM. Statistical comparisons were performed using Mann-Whitney.

Table S7. HLA binding analysis of the 49 class II epitopes identified in 3 or more donors. Related to Fig. 3.

Peptide	DRB1* 01:01	DRB1* 03:01	DRB1* 04:01	DRB1* 04:05	DRB1* 07:01	DRB1* 08:02	DRB1* 09:01	DRB1* 11:01	DRB1* 12:01	DRB1* 13:02	DRB1* 15:01	DRB3* 01:01	DRB3* 02:02	DRB4* 01:01	DRB5* 01:01
CEFQFCNDPFLGVYY	6946	26787	24553	8931	23604	>40k	5606	>40k	>40k	>40k	7576	283	>40k	>40k	7841
SSANNCTFEYVSQPF	3782	26450	8974	3483	519	5263	208	>40k	6177	>40k	663	1070	>40k	7365	4524
CTFEYVSQPFLMDLE	31	5245	11014	115	15	1645	12	30036	2487	>40k	163	92	8958	1291	285
LMDLEGKQGNFKNLR	650	>40k	>40k	1870	2099	12870	7339	30972	>40k	>40k	5036	23143	>40k	>40k	195
NIDGYFKIYSKHTPI	169	>40k	7198	1101	9.7	17	95	28	4077	558	10	547	4459	10230	86
NLVRDLPQGFSALEP	1554	942	28349	19268	1590	7264	7967	1600	1178	247	2438	6.6	9938	6202	6125
RFASVYAWNRKRISN	505	19059	2682	3859	205	59	637	21	11511	6785	92	218	3070	11097	0.86
NLLLQYGSFCTQLNR	705	>40k	3674	99	4816	10477	313	28362	5681	>40k	1.1	7523	>40k	840	3316
NFSQILPDPSKPSKR	1237	1501	256	3016	>40k	14753	24931	1390	>40k	13341	61	204	769	162	2365
AGFIKQYGDCLGDIA	300	>40k	20591	5607	15605	8979	>40k	839	>40k	>40k	6	4203	9508	117	94
SWFTALTQHGKEDLK	72	>40k	398	1643	17069	2261	2078	1117	>40k	>40k	12275	7011	24709	8103	34
DDQIGYYRRATRRIR	111	>40k	10005	5753	215	27	182	15	>40k	>40k	145	553	11802	10210	34
YYRRATRRIRGGDGK	299	>40k	16817	26007	429	36	1314	119	>40k	38689	1688	8659	>40k	9047	57
NKDGIIWVATEGALN	174	24475	792	2140	106	14186	94	>40k	28295	5166	452	9774	22404	812	7085
AGNGGDAALALLLLD	280	>40k	13223	1150	1508	>40k	897	>40k	732	>40k	101	912	>40k	190	4787
DAALALLLLDRLNQL	203	1382	13990	262	4770	809	4173	486	137	593	71	61	>40k	19	6832
LLLLDRLNQLESKMS	505	1572	9927	2469	11692	661	20232	164	453	1048	176	212	3416	143	1562
AAEASKKPRQKRTAT	>40k	>40k	>40k	>40k	>40k	34741	>40k	8920	20999	>40k	17038	>40k	>40k	14790	2655
KKPRQKRTATKAYNV	10628	>40k	>40k	>40k	96	1902	4167	10895	>40k	>40k	9299	>40k	>40k	>40k	6646
KRTATKAYNVTQAFG	2852	>40k	22865	11070	132	12626	447	20127	>40k	7725	1638	11661	>40k	3986	1558
WPQIAQFAPSASAFF	23	>40k	258	492	9.3	395	>40k	2916	>40k	>40k	9.1	66	>40k	>40k	374
PSGTWLTYTGAIKLD	41	6252	7087	6304	6.2	3952	>40k	5860	1015	>40k	100	113	>40k	>40k	30
FKDQVILLNKHIDAY	99	4799	13537	695	414	148	>40k	433	148	>40k	81	2248	>40k	37936	2754
ILLNKHIDAYKTFPP	627	>40k	>40k	8357	3212	1254	>40k	1974	>40k	>40k	14	6769	>40k	>40k	1094
TQHQPYVVDDPCPIH	22686	21176	>40k	>40k	17613	>40k	35313	>40k	>40k	32533	8335	79	>40k	19493	30768
YVVDDPCPIHFYSKW	>40k	803	9265	>40k	>40k	27613	2768	20527	>40k	25989	656	58	27684	906	>40k
PCPIHFYSKWYIRVG	6513	21090	4597	900	239	1220	753	31597	>40k	24315	21	22344	>40k	630	1315
FYSKWYIRVGARKSA	1.8	38596	498	179	8.6	25	55	54	5727	>40k	164	422	5253	2613	0.77
IGNYTVSCLPFTINC	92	>40k	3302	4098	44	3771	101	2333	4720	20150	1994	612	11780	>40k	360
FTINCQEPKLGSLVV	506	>40k	>40k	>40k	3294	3898	5936	920	529	16253	4990	5352	1705	98	4087
GSLVVRCSFYEDFLE	8214	4608	1331	309	964	10509	290	>40k	19312	>40k	206	83	>40k	>40k	8478
RCSFYEDFLEYHDVR	484	9278	36490	5172	1988	735	3586	4322	25468	>40k	3142	1042	36663	6053	210
QFAYANRNRFLYIIK	11	1138	3944	769	120	93	244	117	72	154	33	193	199	19536	54
VLAAVYRINWITGGI	121	36604	351	252	74	151	38	92	10356	455	91	13953	78	487	92
YRINWITGGIAIAMA	9.3	>40k	7073	4895	12	351	111	1393	3543	2624	13	4116	631	633	1543
CLVGLMWLSYFIASF	505	>40k	31869	3646	5859	>40k	339	30074	4438	>40k	90	27076	>40k	>40k	11097
MWLSYFIASFRLFAR	380	10769	5192	771	262	119	111	33	27091	>40k	87	20263	>40k	990	131
TNILLNVPLHGTILT	55	4228	694	909	1031	528	404	1010	553	33	128	946	119	997	3466
SELVIGAVILRGHLR	46	34557	20482	8846	611	12106	1430	25154	790	1249	604	8759	6162	2551	138
RGHLRIAGHHLGRCD	64	18371	571	7860	235	55	92	82	>40k	3954	35	1843	2799	55	67
IAGHHLGRCDIKDLP	2134	>40k	>40k	>40k	8926	>40k	>40k	37099	2575	21846	2187	>40k	>40k	>40k	15863
LGRCDIKDLPKEITV	>40k	>40k	>40k	>40k	>40k	3514	>40k	3900	7059	7949	28038	2237	>40k	27485	13289
IKDLPKEITVATSRT	2644	>40k	7673	7553	5189	4601	380	6836	>40k	33962	6817	>40k	>40k	1030	2392
KEITVATSRTLSYYK	2.4	208	929	8454	19	283	24	229	2924	246	5.1	2761	14511	504	42
LSYYKLGASQRVAGD	0.85	14589	88	178	13	279	5.7	747	>40k	4589	283	82	51	9484	2.5
SGFAAYSRYRIGNYK	138	>40k	7026	4143	620	37	71	259	>40k	>40k	63	529	7025	36269	31
LYLYALVYFLQSINF	154	>40k	2945	368	91	24356	91	>40k	25	>40k	76	1957	28791	242	2947
QSINFVRIIMRLWLC	139	9575	3335	512	268	546	2170	497	89	>40k	16	>40k	24678	164	31
LMIERFVSLAIDAYP	12	6590	288	76	8.1	670	3.8	1916	16	5367	0.084	262	961	1.4	298

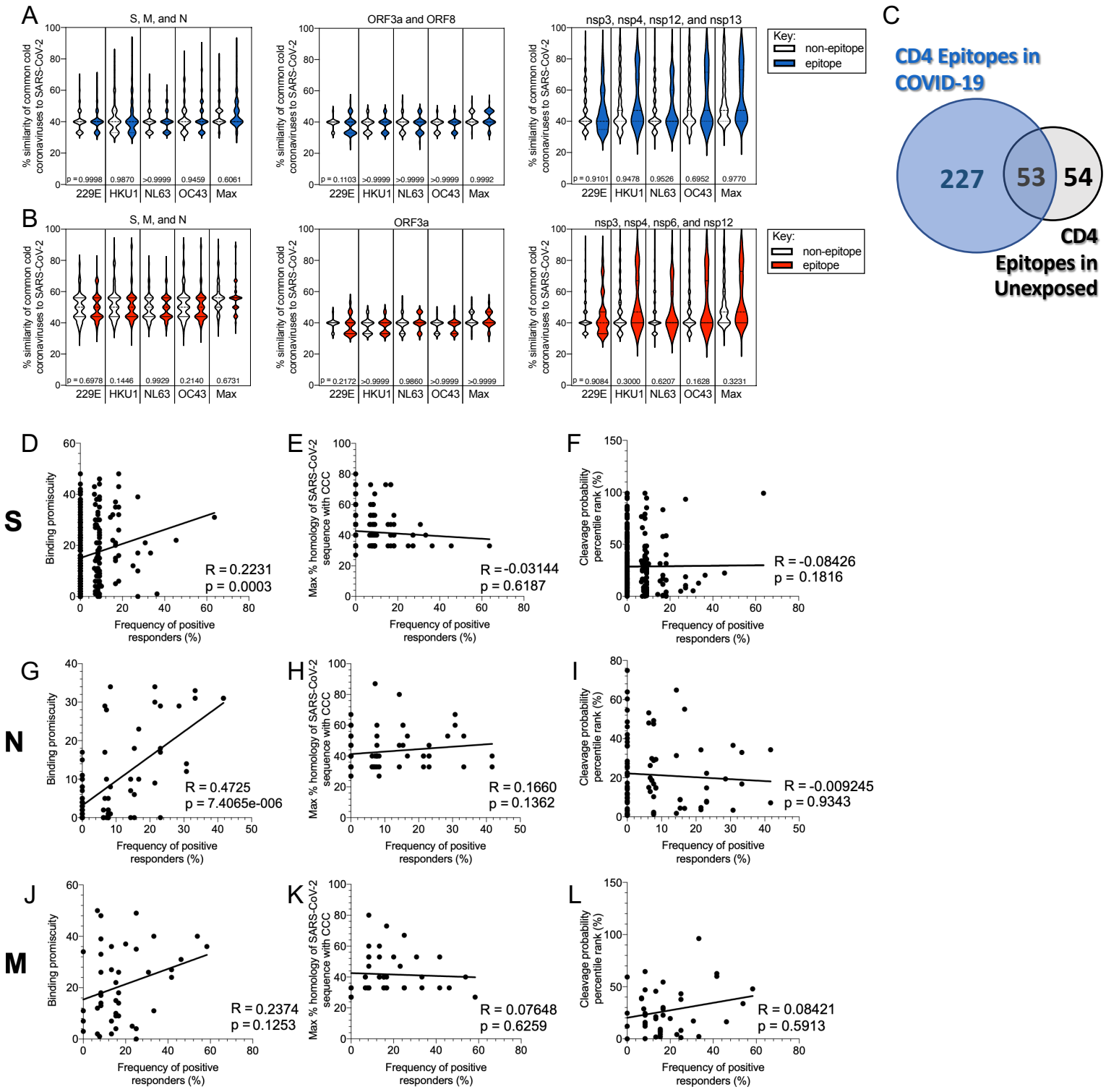

Fig. S5. **Analyses of CD4⁺ and CD8⁺ T cell epitopes identified compared to non-epitopes and correlations of binding promiscuity, homology to CCC, and frequency of positive response. Related to Fig. 4-5.**

Comparison of sequenced identity between CD4⁺ T cell epitopes and non-epitopes as a function of sequence identity with the CCC in S, M, and N combined, ORF8 and ORF3a, and non-structural proteins (**A**). For CD8⁺ epitopes and non-epitopes, the sequence identities with CCC are shown for S, M, and N, ORF3a, and non-structural proteins (**B**). Statistical analyses were performed using the Kolmogorov-Smirnov test, and data are shown as violin plots. (**C**) Overlap of previously identified epitopes in unexposed (Mateus et al., 2020 Science) with the proteins analyzed in this study and the current epitopes identified in COVID-19 donors. The Venn diagram was calculated with the Venn Diagram Plotter (PNNL, OMICS.PNL.gov). Correlations of predicted binding promiscuity to the alleles present in the donor cohort tested with the frequency of positive response for S (**D**), N (**G**), and M (**J**) epitopes. Frequency of positive response is also correlated with the maximum % homology of the SARS-CoV-2 sequence to CCC and plotted for S (**E**), N (**H**), and M (**K**). In the final column of panels, the correlation of frequency of positivity and the cleavage probability percentile rank (calculated using the IEDB MHCII-NP tool) are shown for S (**F**), N (**I**), and M (**L**). Statistics were performed using the Spearman correlation and the line on each graph is a simple linear regression.

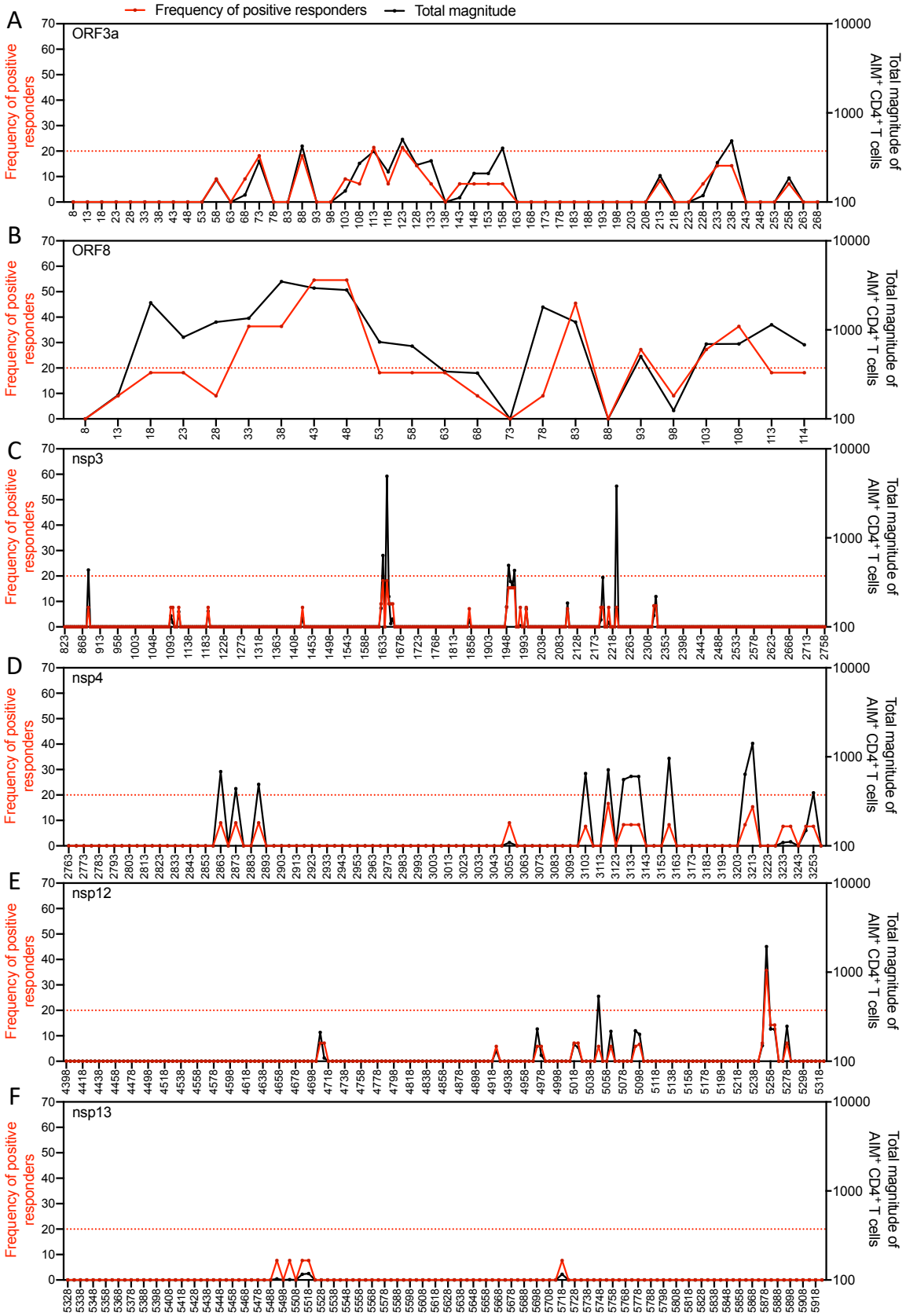

Fig. S6. Immunodominant regions for the other major antigens for CD4⁺ T cells. Related to Fig. 5.
Immunodominant regions for the other major antigens for CD4⁺ T cells: ORF3a (**A**), ORF8 (**B**), nsp3 (**C**), nsp4 (**D**), nsp12 (**E**), and nsp13 (**F**). The frequency of positive responders is shown in red and the total magnitude of response (sum of AIM⁺ T cells for each peptide) is shown in black. The x-axis is labeled with the middle, or 8th, position of each 15-mer peptide. The dotted red line at 20% frequency of positive responders indicates a cutoff for immunodominant epitopes.