Cell Reports, Volume *34*

Supplemental information

Linear epitope landscape of the SARS-CoV-2 Spike protein constructed from 1,051 COVID-19 patients

Yang Li, Ming-liang Ma, Qing Lei, Feng Wang, Wei Hong, Dan-yun Lai, Hongyan Hou, Zhao-wei Xu, Bo Zhang, Hong Chen, Caizheng Yu, Jun-biao Xue, Yun-xiao Zheng, Xue-ning Wang, He-wei Jiang, Hai-nan Zhang, Huan Qi, Shu-juan Guo, Yandi Zhang, Xiaosong Lin, Zongjie Yao, Jiaoxiang Wu, Huiming Sheng, Yanan Zhang, Hongping Wei, Ziyong Sun, Xionglin Fan, and Sheng-ce Tao

Linear epitope landscape of the SARS-CoV-2 Spike protein constructed from 1,051 COVID-19 patients

Li *et al.*

Supplemental figures

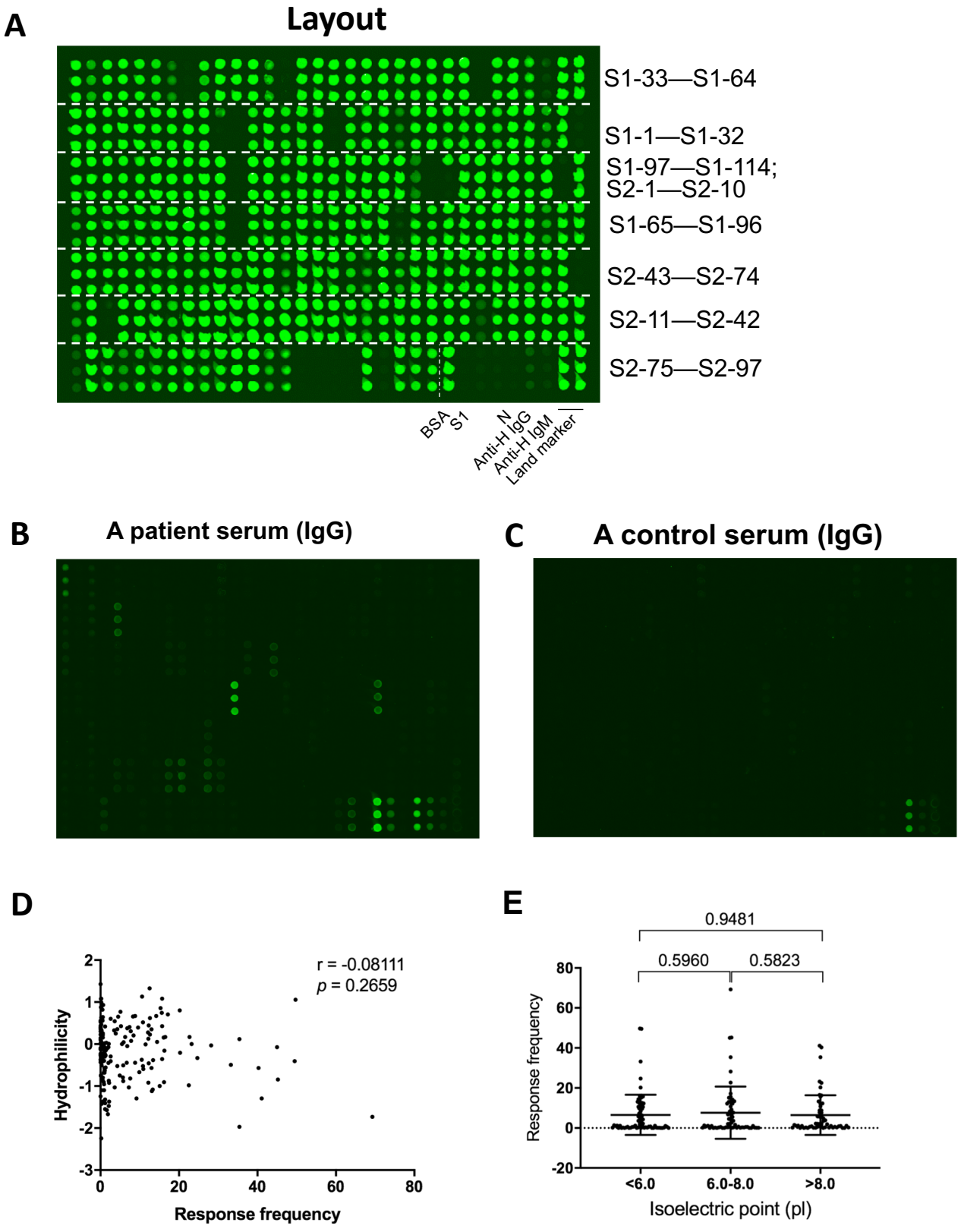


Figure S1. The peptide microarray (related to Figure 1 and Table S1). A. The layout of the peptide microarray that was used in this study (**Table S1**). **B**. An example array probed with COVID-19 patient serum. **C**. An example array probed with a control serum. **D**. Correlations between the hydrophilicity and the response frequency of the epitopes. Each spot indicates one peptide. The P value was calculated with the two-sided F-test. **E**. Statistical analysis of the response frequency of the peptides with low, medium and high pI values. The P value was calculated with the two-sided t-test.

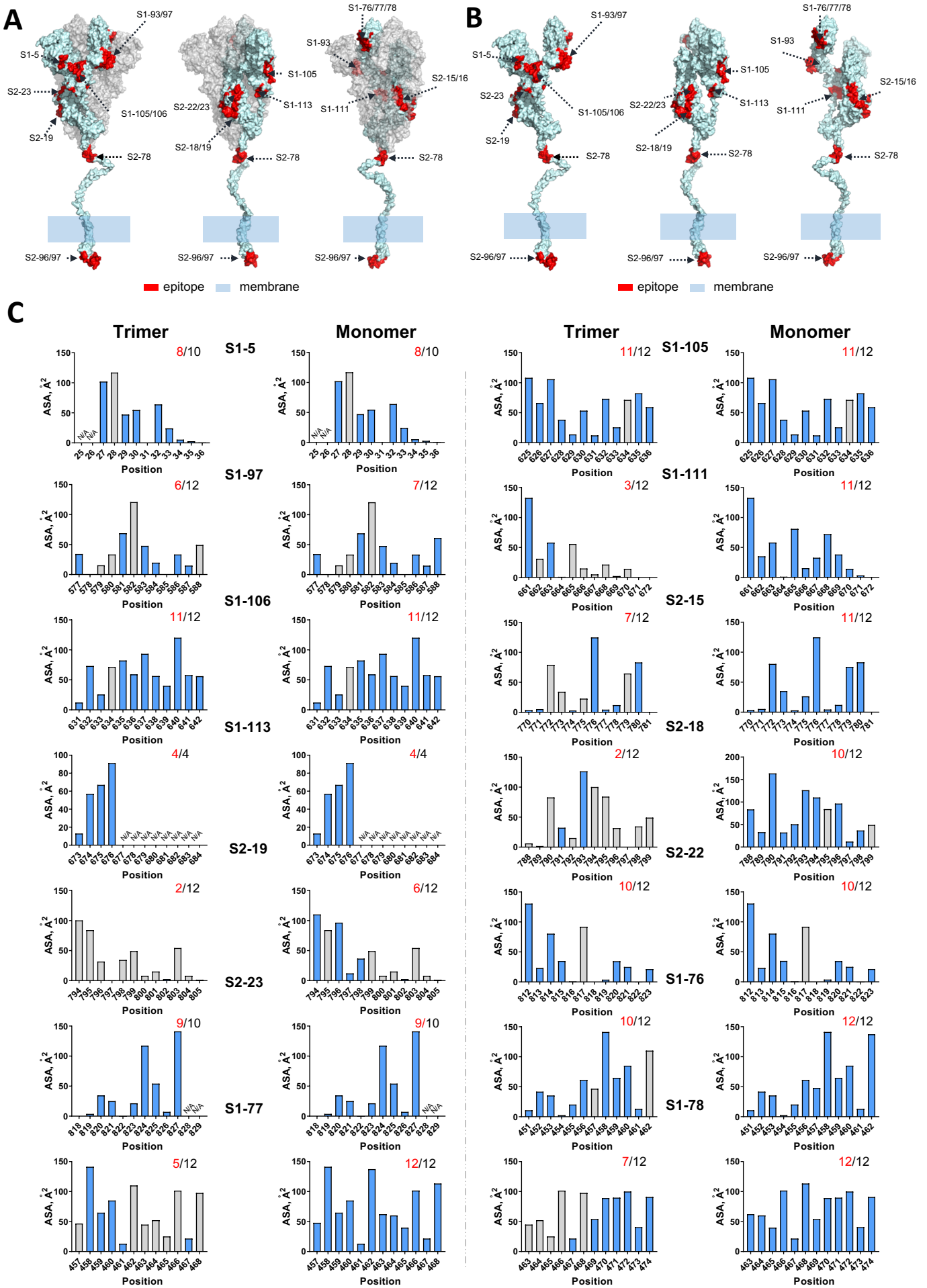


Figure S2. The distribution of the highly immunogenic epitopes on the Spike protein (related to Figure 1)

Figure S2. The distribution of the highly immunogenic epitopes on the Spike protein (related to Figure 1). A-B. **A**. The 3D structure of the spike protein is used (PDB ID: 6X6P). The section from S2-78 to the end of the C-terminus was modelled using C-I-TASSER. The 19 significant epitopes are marked in red on the 3D structure of the Spike protein for both the trimer (**A**) and the monomer (**B**). **C**. The area of solvent accessibility (ASA) of each amino acid on the indicated epitope for trimer and monomer format concerning an S protein trimer structure (PDB: 6X6P).

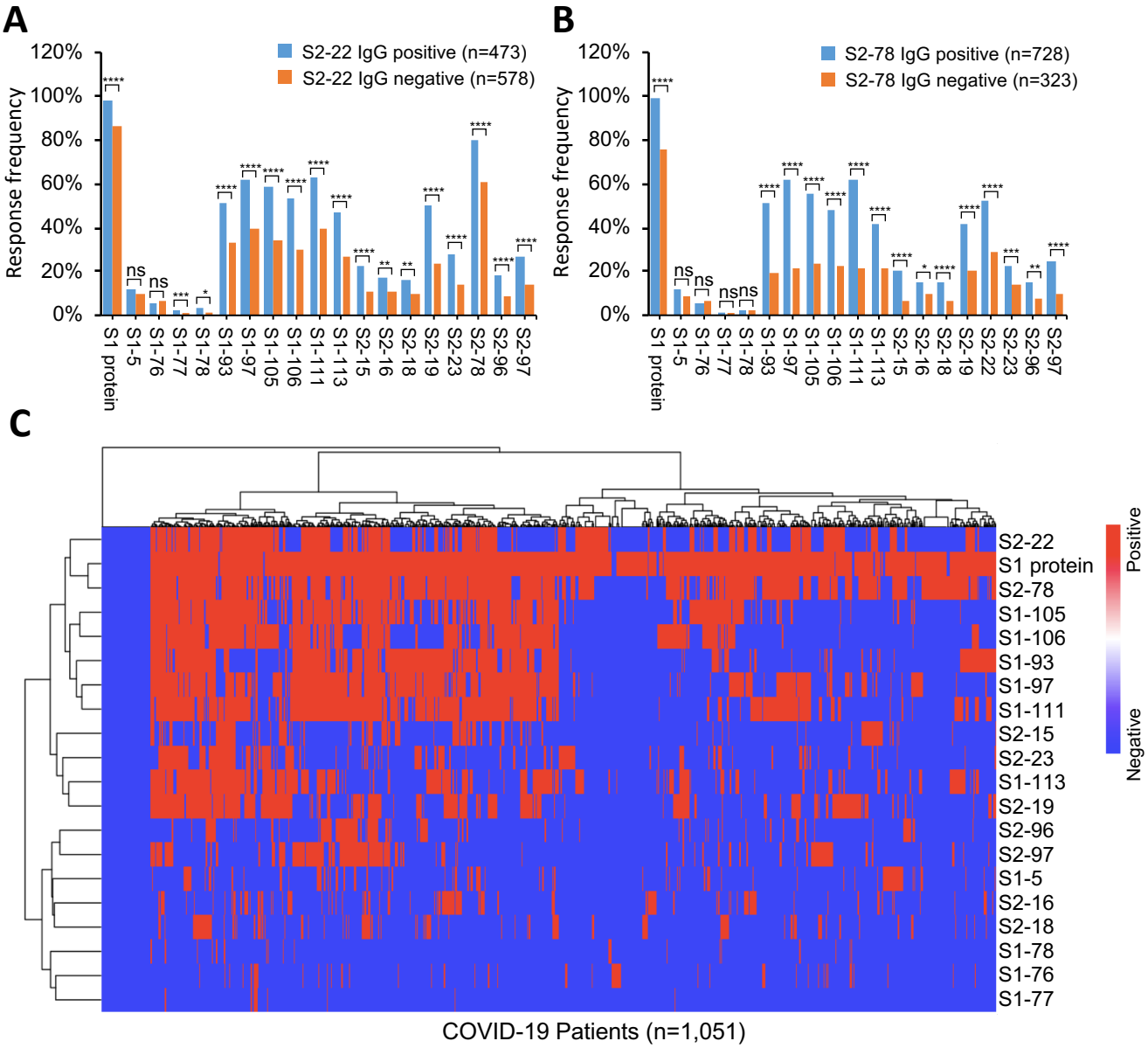


Figure S3. IgG response signatures against Spike liner epitopes in COVID-19 patients (related to Figure 1). **A-B.** The response frequency for each epitope in the two groups as S2-22 (**A**) or S2-78 (**B**) IgG positive or negative. The *P* value was calculated with the χ^2 test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant. **C** Heatmap of clustering analysis IgG response signatures for COVID-19 patients. Each patch indicates positive (red) or negative (blue) IgG response against the significant epitope or S1 protein (row) in one patient (column).

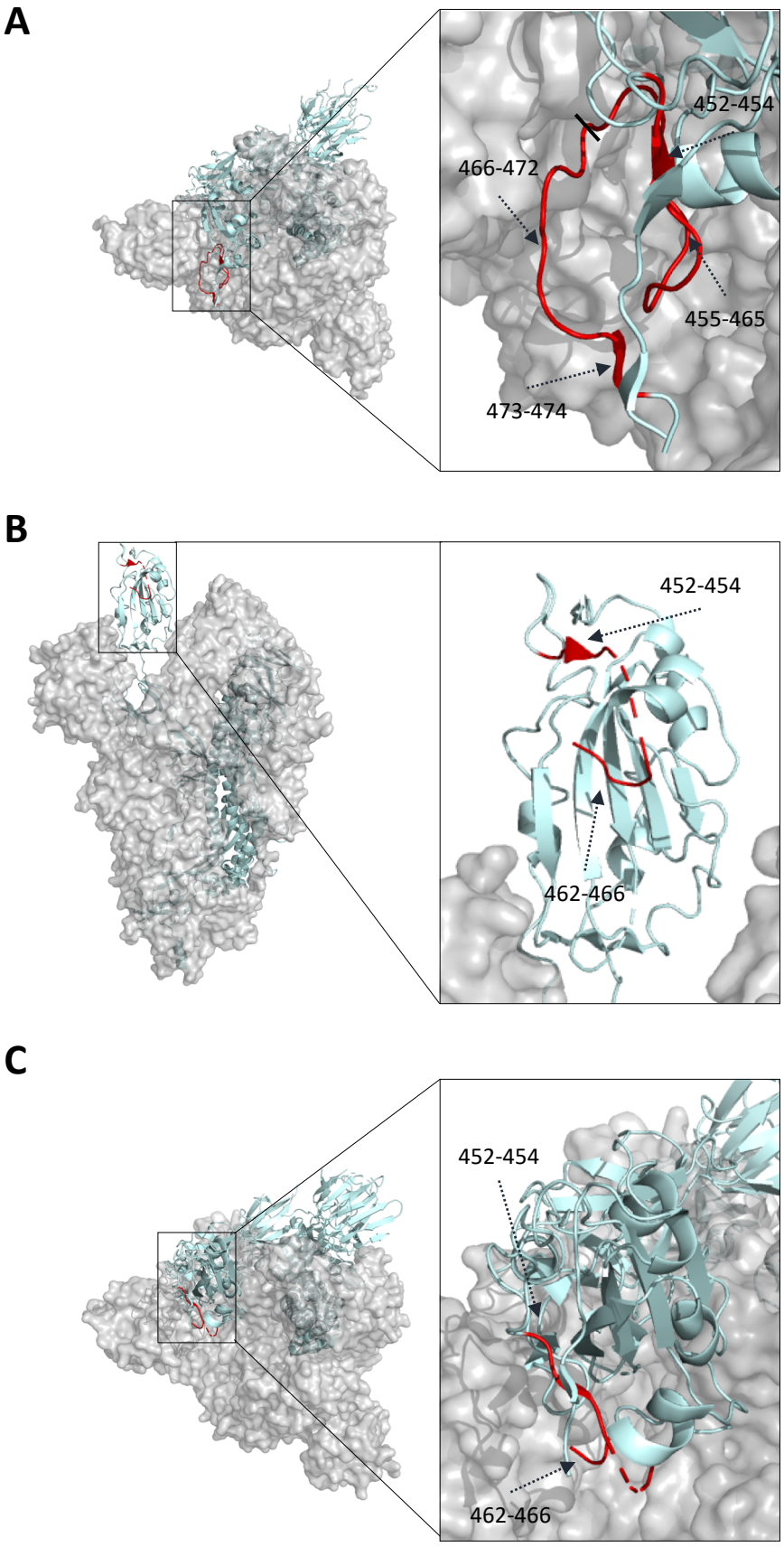
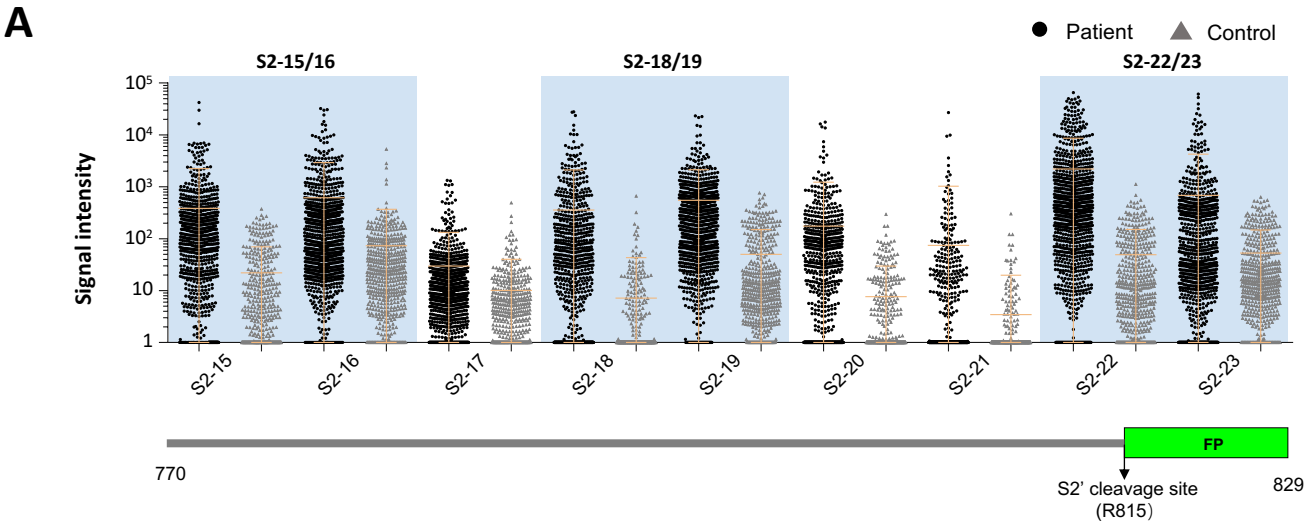
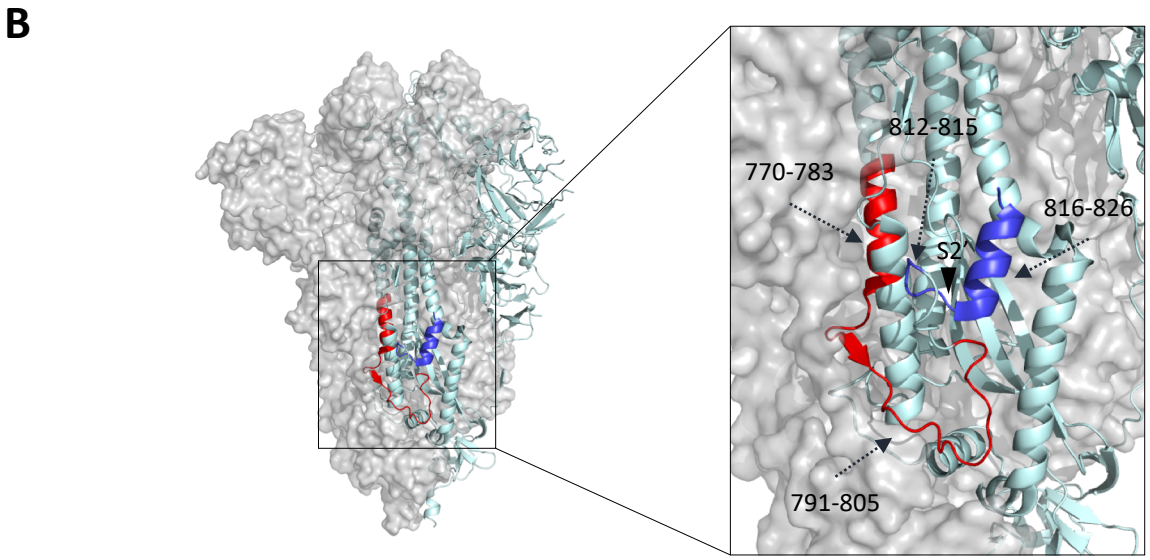


Figure S4. The location of S1-76/77/78 on the RBD (related to Figure 4). A. A top-down view of the closed-state Spike protein trimer (PDB: 6X6P). **B.** A side view of the open-state Spike protein trimer (PDB ID: 6VYB). **C.** A top-down view of the open-state Spike protein trimer (PDB ID: 6VYB). The significant epitopes (S1-76/77/78, aa451-474) are marked in red.





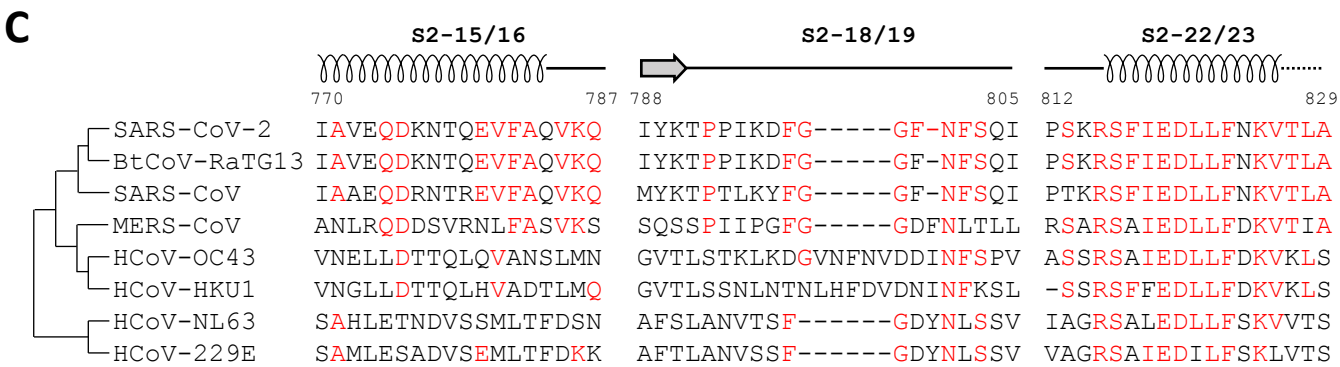
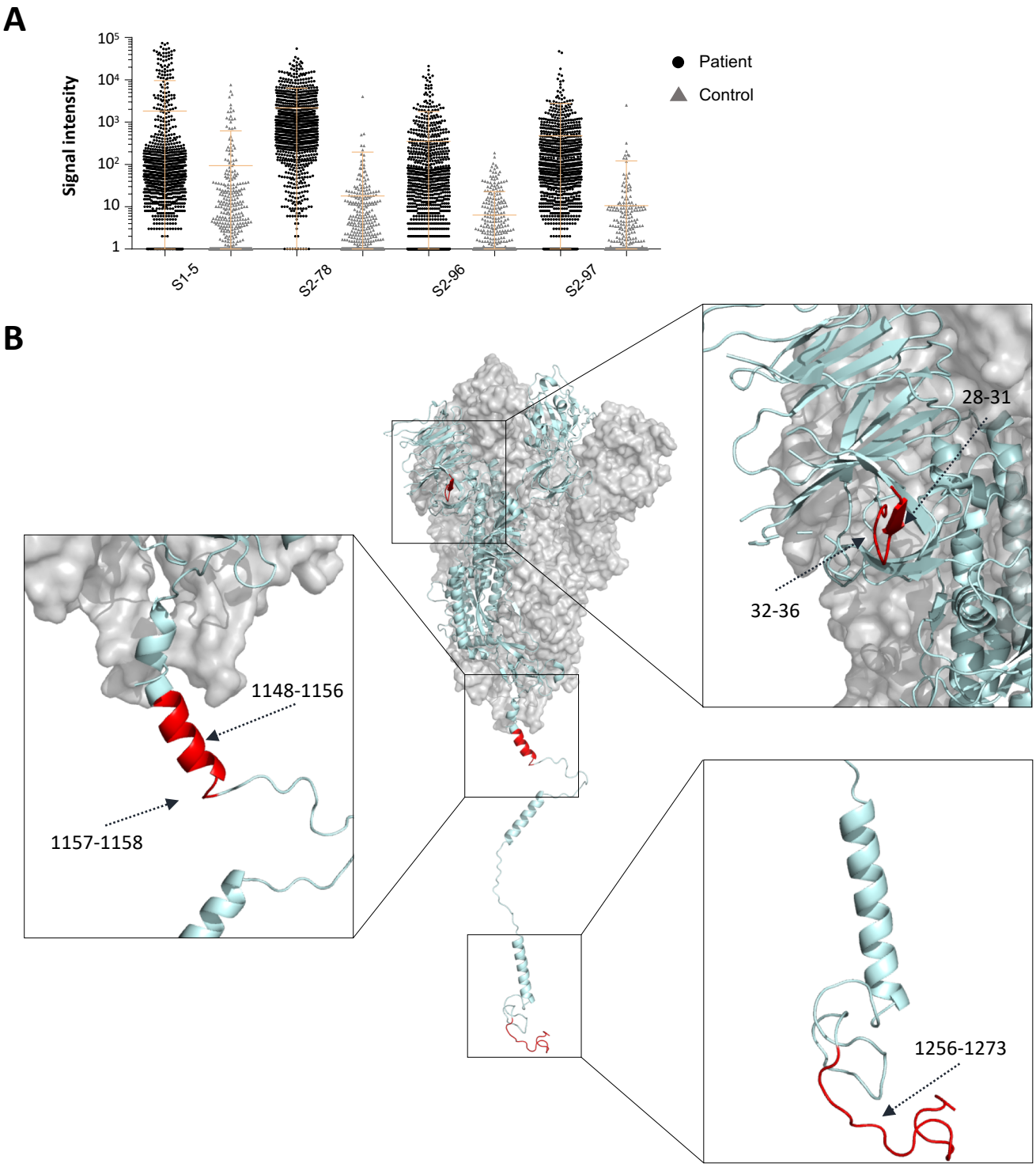


Figure S5. The 2nd hot spot of highly immunogenic linear epitopes: S2'cleavage site and FP (related to Figure 5). A. The S2'cleavage site and FP in the linear epitope landscape. **B**. The significant epitopes are located in this region. S2-15/16, aa770-787, red, coil; S2-18/19, aa788-805, red, loop; and S2-22/23, aa812-829, blue, coil. **C**. The homology analysis of the significant epitopes among the 7 known human coronaviruses and bat coronavirus BtCoV-RaTG13. The amino acids with consistencies >=50% among the 8 coronaviruses are marked in red. The loop, α-helix and β-strand region are shown as a line, a coil and an arrow above the sequences, respectively. An unobserved structure is shown as a dotted line.



C

	S1-5	S2-78	S2-96/97
	25 36	1148 1159	1256 1273
SARS-CoV-2	P---PAY--TNSFTRGV	FKEELDKYFKNH	F-DEDDSEPVLKGVKLHYT
BtCoV-RaTG13	P---PAY--TNSSTRGV	FKEELDKYFKNH	F-DEDDSEPVLKGVKLHYT
SARS-CoV	A---PNYTQHTSSMRGV	FKEELDKYFKNH	F-DEDDSEPVLKGVKLHYT
MERS-CoV	W---PRP-IDVSKADGI	FQDELDEFFKNV	RYEEYDLEPHKVHVH----
HCoV-OC43	PPPISTDTVDVTNGLGT	FKEELDQWFKNQ	DDYTGYQELVIK-TSHDD-
HCoV-HKU1	VPRISEYVVDVSYGLGT	FESELSHWFKNQ	DEYGGHHDFVIK-TSHDD-
HCoV-NL63	WFC-ANQSTSVYSANGF	VNKTLQEFAQNL	PYYEFEKVHVQ--------
HCoV-229E	----------	VNKTLQELSYKL	PYYDVEKIHIQ--------

Figure S6. Other highly immunogenic linear epitopes (related to Figure 5)

Figure S6. Other highly immunogenic linear epitopes (related to Figure 5). A. An additional 5 significant epitopes that do not belong to the two "hot spots". **B**. The significant epitopes are located on the Spike protein. S1-5, aa25-36, red; S2-78, aa1148-1159, red; and S2-96/97, aa1256-1273, red. **C**. The homology analysis of the significant epitopes among the 7 known human coronaviruses and the bat coronavirus BtCoV-RaTG13. The amino acids with consistencies >=50% among the 8 coronaviruses are marked in red. The loop, α-helix and β-strand region are shown as a line, a coil and an arrow above the sequences, respectively. An unobserved structure is shown as a dotted line.

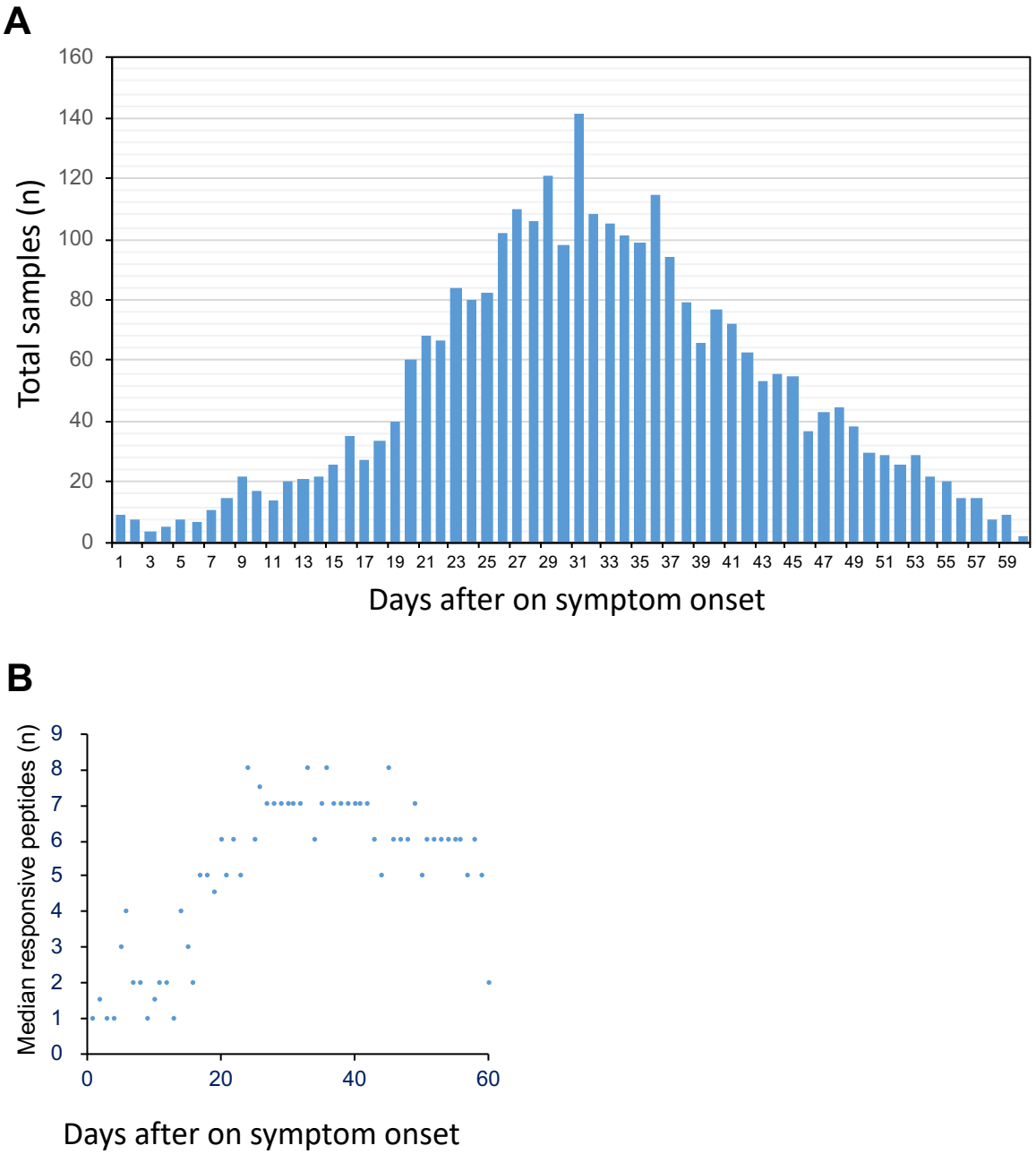


Figure S7. **Dynamic changes in responsive** epitope **numbers (related to Figure 6)**. **A**. The number of serum samples for each day. **B**. The median number of responsive peptides for the samples collected at the indicated time point.