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

Yang Li[1+], Ming-liang Ma[1+], Qing Lei[2+], Feng Wang[3+], Dan-yun Lai[1], Hongyan Hou[3], Zhao-wei Xu[1], Bo Zhang[3], Hong Chen[1], Caizheng Yu[4], Jun-biao Xue[1], Yun-xiao Zheng[1], Xue-ning Wang[1], He-wei Jiang[1], Hai-nan Zhang[1], Huan Qi[1], Shu-juan Guo[1], Yandi Zhang[2], Xiaosong Lin[2], Zongjie Yao[2], Jiaoxiang Wu[5], Huiming Sheng[5], Ziyong Sun[3, *], Xionglin Fan[2, *], Sheng-ce Tao[1, *]

[1]Shanghai Center for Systems Biomedicine, Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Jiao Tong University, Shanghai, China

[2]Department of Pathogen Biology, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Clinical Laboratory, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[4]Department of Public Health, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

[5]Tongren Hospital, Shanghai Jiao Tong University School of Medicine，Shanghai, China

[+] These authors contribute equally to this work.

[*]Corresponding Authors, e-mail: taosc@sjtu.edu.cn (S.-C. Tao), xlfan@hust.edu.cn (X.-L. Fan), zysun@tjh.tjmu.edu.cn (Z.-Y. Sun)

## Abstract

Neutralization antibodies and vaccines for treating COVID-19 are desperately needed. For precise development of antibodies and vaccines, the key is to understand which part of SARS-CoV-2 Spike protein is highly immunogenic on a systematic way. We generate a linear epitope landscape of Spike protein by analyzing serum IgG response of 1,051 COVID-19 patients with a peptide microarray. We reveal two regions that rich of linear epitopes, *i.e.*, CTD and a region close to the S2' cleavage site and fusion peptide. Unexpectedly, we find RBD is lack of linear epitope. Besides 3 moderate immunogenic peptides from RBD, 16 highly immunogenic peptides from other regions of Spike protein are determined. These peptides could serve as the base for precise development of antibodies and vaccines for COVID-19 on a systematic level.

## One sentence summary

A linear epitope landscape of SARS-CoV-2 Spike protein is generated by analyzing serum IgG response of 1,051 COVID-19 patients.

COVID-19 is caused by SARS-CoV-2 (*1*, *2*) and is a global pandemic. By July 12, 2020, 12,728,966 cases were diagnosed and 565,351 lives were claimed (https://coronavirus.jhu.edu/map.html) (*3*). Highly effective neutralization (prophylactic or therapeutic) antibodies and vaccines are pressingly needed. The genome of SARS-CoV-2 encodes 27 proteins, among them, Spike protein plays a central role for the binding and entry of the virus to the host cell. Spike protein is cleaved into S1 and S2 at furin and S2' sites by specific proteases (*4*). It is highly glycosylated with 21 N-glycosylation sites (*5*). Spike protein and more specific, the RBD domain is currently the most-focused target for the development of COVID-19 neutralization antibodies and vaccines (*6–11*). However, except RBD, other parts on Spike protein may also elicit neutralization antibodies (*12–16*). In addition, antibody dependent enhancement (ADE) is an unneglectable concern for the development of neutralization antibody and vaccine for SARS-CoV-2 (*17*, *18*). To facilitate the precise development of neutralization antibodies and vaccines, it is necessary to understand which part of Spike protein is highly immunogenic on a systematic way, by analyzing samples collected from a large cohort of COVID-19 patients.

## Results

### The IgG linear epitope landscape of SARS-CoV-2 Spike protein

To reveal the immunogenic linear epitopes of Spike protein, a peptide microarray with full coverage of Spike protein was updated from an original version (*14*) (**Fig. S1**). Sera were collected from two groups, *i.e.*, 1,051 COVID-19 patients and 528 controls (**Table S1**), and analyzed individually on the peptide microarray (**Table S2)**. By plotting the signal intensities of all the samples against each peptide, a linear epitope landscape was constructed, for better overview, the landscape is aligned to the sequence of Spike protein (**Figure 1**). To assure the specificity, all the control samples (**Table S1**) were also analyzed on the peptide microarray. Almost all the peptides were negative for all the control samples, while significant bindings were observed for many of the peptides when probed with COVID-19 sera. This indicate that the positive bindings are SARS-CoV-2 specific. It was speculated that the N-glycosylation may interfere with antibody responses (*19*). However,

our results clearly showed that the distribution of the linear epitopes is not related to N-glycosylation, if any, subtle.

To call the highly immunogenic peptides, the criteria were set as the average_signal intensity is larger than 3*Cutoff2, and the response frequency is larger than 10% (see methods for the definitions). A total of 16 peptides were obtained, surprisingly, all of them are outside of RBD. Because the significance of RBD, we lowered the criteria, *i. e.*, the response frequency is larger than 1%, while keep the average_signal intensity larger than 3*Cutoff2, 3 consecutive peptides of moderate immunogenicity, *i. e*, S1-76, S1-77 and S1-78 were selected (**Table S3)**. Interestingly, all these peptides locate within RBM (receptor binding motif), the binding interface of Spike protein and ACE2.

While a few of these immunogenic peptides are dispersed on Spike protein, there are two linear epitope "hot" regions could be immediately recognized, aa525-685 and aa770-829, one of which is CTD (C-terminal domain) and another covers the S2' cleavage site and the fusion peptide (FP). For the immunogenic peptides (**Table S3),** the solubility (*20*) and pI range from -1.97 to 1.06, and 3.01 to 11.16, respectively. This indicates the overall immunogenicity of Spike protein is not correlated to solubility and pI. There are several SARS-CoV-2 epitope related studies involving small sample sets (*21*, *22*). Our immunogenic epitopes are partially consistent with these studies (**Table S3**). A relative high consistency was observed between our data and ReScan, a phage display based strategy (*23*, *24*).

It is known that the immune response may be related to some key clinical parameters, such as gender, disease severity, age and the final outcome (*25*, *26*). To test whether the linear epitope landscape correlates with these parameters, we divided the linear landscape (**Fig. 1A**) according to these parameters, *i.e.*, male vs. female for gender (**Fig. S2**.), mild vs. severe for severity (**Fig. S3**), >60 vs. <60 for age (**Fig. S4**) and recovered vs. death for final outcome (**Fig. S5**). Unexpectedly, overall, no obvious difference was observed. These suggest that the linear epitope landscape is highly robust and consistent among COVID-19 patients.

To further illustrate the location and distribution of the 19 immunogenic peptides (**Table S3**), we mapped them to the 3D structure of Spike protein (*27*) (**Fig. S6A** and **S6B**). It is clear that most of these peptides locate on the surface of Spike protein, which is consistent

to the common notion (*28*). However, S1-111, S2-15 and S2-16 are not on the surface of the trimeric Spike protein, but on the surface of the monomer. A plausible explanation is that the monomer Spike protein could be exposed to the immune system at a certain yet undiscovered stage.

**RBD is lack of linear epitope while highly immunogenic**

It is well known that RBD is highly immunogenic (*26*, *29*). However, according to our selection criteria, no highly immunogenic peptide was obtained from RBD. Only when we lowered the selection criteria, 3 peptides were selected, *i.e.*, S1-76, S1-77 and S1-78 (**Fig. 2A**). When RBD is compared to other regions of Spike protein, it is obvious that RBD is very poor of linear epitopes. This seems contradictory to the knowledge that RBD is highly immunogenic. It is possible that most of the epitopes of RBD region are conformational. To test this possibility, we collected a set of 9 high affinity monoclonal antibodies of RBD or Spike protein (see methods). These antibodies were obtained through memory B cell isolation from COVID-19 recovered patients (*30*). We analyzed these antibodies individually on the Spike protein peptide microarray (*14*) (**Fig. 2B**). Among these antibodies, 414-1 has the highest affinity (2.96 nM) to RBD, as expected, strong bindings were observed for both S1 protein and RBD, however, negative signals were obtained for all the peptides, including the RBD peptides from aa331-524. Also, no peptide bindings were observed for the rest of the RBD specific antibodies (data not shown). For antibody 414-4, strong binding was obtained for S1 protein but not RBD. Interestingly, 414-4 binds S1-97 with high affinity, indicating the epitope that 414-4 recognized is around aa577-588. Actually, the 3 immunogenic peptides, *i. e*., S1-76, S1-77 and S1-78 were also identified in other related studies (*22*, *31*). These three peptides are consecutive and locate in RBM region, which are at least partially overlapped with or close to the binding epitopes of a variety of SARS-CoV-2 neutralization antibodies, *e.g*., B38 (*6*), CB6 (*32*) and P2B-2F6 (*33*). To further illustrate the locations of these peptides, we mapped them to the 3D structure of Spike protein of both the open state and the closed state (**Fig. 2C**). In closed state, aa455-465 of S1-76/77/78 locate at the contact area among the three monomers and probably is difficult to be accessed, aa452-454 and 473-474 form as β-strand, and are covered but could be accessed from both sides, while only aa466-472 are exposed and

present as flexible sequence (**Fig. 2C** and **Fig. S7**). In the open state of Spike protein, all residues of S1-76/77/78 are exposed and highly accessible (**Fig. S7**). To further analyze the immunogenicity of S1-76/77/78, neutralization antibody-RBD complexes are used (**Fig. S8**), the antibodies are CB6 (*32*), P2B-2F6 (*33*), BD23 (*7*), CR3022 (*8*) and S309 (*34*). Among these structures, CB6 and P2B-2F6 interact directly with residues within S1-76/77/78. For the CB6-RBD complex (**Fig. S8D**), there are several residues within S1-76/77/78, which have direct interactions with the antibody, *i. e.*, Y453, L455, F456, R457, K458, S459, N460, Y473 and Q474. While for the P2B-2F6-RBD complex (**Fig. 8E**), the only residue which has direct interaction with the antibody is L452.

Broad neutralization antibody and vaccine effective for SARS-CoV-2 and other human coronaviruses are of high interest (*35*). We performed the homology analysis for S1-76/77/78 among SARS-CoV-2, the other 6 human coronaviruses, and bat coronavirus BtCoV-RaTG13 (*1*). High homologies were observed for all these 3 peptides among SARS-Cov-2, SARS-CoV and BtCoV-RaTG13. The high homology indicate that antibodies elicited by S1-76/77/78 may at least be effective for both SARS-Cov-2 and SARS-CoV.

These results strongly suggest that RBD is rich of conformational epitopes while poor of linear epitopes. The underlying mechanism is worth for further investigation. Taken together, our results suggest that S1-76/77/78 could serve as a promising candidate for the development of broadly neutralization antibodies or vaccine.

**CTD is rich of linear epitopes**

The first "hot" region of linear epitopes is CTD. The whole domain is densely covered by linear epitopes (**Fig. 3A**). According to the selection criteria, 6 highly immunogenic epitopes, *i.e.*, S1-93, S1-97, S1-105/106, S1-111 and S1-113 were identified. These peptides are about evenly distributed cross CTD. We then asked whether these 6 highly immunogenic peptides/ epitopes were also revealed by other studies. It showed that S1-93 was identified by ReScan (*23*) （**Table S3**) and COVIDep (*21*), S1-97 by ReScan, and S1-111 by COVIDep.

S1-93 and S1-97 locate at CTD1, aa555-564 of S1-93 and aa578-584 of S1-97 present as loop region and are on the surface of trimeric Spike protein. S1-105, S1-106, S1-111,

S1-113 locate at CTD2, S1-105/106 are almost loop and present on the surface, S1-111 is β-strand and loop but buried inside, only aa667-669 on loop region could be accessed. S1-113 is nearly to S1/S2 cleavage site, although aa677-684 is invisible in Spike protein structure, these residues could be exposed on surface and induce antibody response to prevent S1/S2 hydrolysis (**Fig. 3B**). S1-113 is also on the out surface while S1-111 is on the inner surface, the underlying mechanism why S1-111 triggers strong IgG reaction in many patients worth further study.

Homology analysis was performed for the 6 highly immunogenic peptides (**Fig. 3C**). Except S1-113, high homologies were observed for all the 5 peptides among SARS-Cov-2, SARS-CoV and BtCoV-RaTG13. The high homology indicate that antibodies elicited by S1-93, S1-97, S1-105 and S1-111 may be effective for both SARS-Cov-2 and SARS-CoV. And antibody targeting S1-113 may specific for SARS-CoV-2.

D614G mutant is the current dominant strain in Europe (*36*), which has about 9 times higher infection efficiency in cell assay than that of the wild type strain (*37*). D614 is within S1-102, a peptide of moderate immunogenicity, and close to the highly immunogenic peptide S1-105. Block the D614 region may cause functionally significant effect.

**The second epitope hot region spans aa770-829, covers the S2' cleavage site and FP**

The second region with highly enriched linear epitopes spans aa770-829 (**Fig. 4A**). According to the selection criteria, 6 highly immunogenic peptides were obtained, *i.e.*, S2-15, S2-16, S2-18, S2-19, S2-22 and S2-23.

It is interesting to see whether these 6 highly immunogenic peptides/ epitopes were also identified in related studies. We revealed that S2-22 was identified by a peptide microarray study (*22*). Four peptides (S2-18, S2-19, S2-22 and S2-23) were identified by ReScan (*23*), and 2 peptides (S2-22 and S2-23) were predicted by COVIDep (*21*). Of these peptides, S2-22 is the only one that was identified or predicted in all these studies. Since S2-22/23 covers the S2' cleavage site and FP, we speculate that antibody targeting S2-22/23 may block the cleavage and disturb the function of FP, thus has potent neutralization activity. Interestingly, strong S2-22 specific IgG reaction was also elicited by a mRNA vaccine study (*38*), which further demonstrated the high immunogenicity and high potential of neutralization activity of S2-22.

To further illustrate the locations of these peptides, we mapped them to the 3D structure of Spike protein. S2-15, S2-16, S2-18, S2-19, S2-22 and S2-23 all locate near S2' site and the FP region. aa770-783 of S2-15/16 form as α-helix and are buried at the trimer interface, but accessible on monomer. aa791-805 of S2-18/19 form as loop and aa816-826 form as α-helix and locate on surface, the S2' cleavage site is on S2-22 (**Fig. 4B**).

To check the similarity of the peptides among human coronaviruses, we performed the homology analysis for S2-22/23, S2-15/16 and S2-18/19. High homologies were observed for all these peptides among SARS-Cov-2, SARS-CoV and BtCoV-RaTG13 (**Fig. 4C**). Interestingly, S2-22/23 is highly homologous among all the coronaviruses, and almost identical among SARS-Cov-2, SARS-CoV and BtCoV-RaTG13. If this peptide can elicit strong neutralization activity, it may could serve as a promising candidate for making broad neutralization antibody and vaccine.

**Other highly immunogenic linear epitopes**

Except the immunogenic peptides that belong to RBD and the two "hot" regions, there are also other highly immunogenic peptides dispersed across Spike protein (**Fig. 5A**). S1-5 locates at NTD, part of the residues, *i. e*., aa28-31 form β-strand and are on the surface of trimeric Spike protein, aa32-36 form as loop and are partially covered by other region. S2-78, S2-96 and S2-97 locate at unobserved regions in the C-terminal of Spike protein. We applied a modeling structure to present these unobserved regions (**Fig. 5B**). S2-78 is predicted as α-helix and S2-96/97 is predicted as loop. S2-96/97 are at the very C-terminal end of Spike protein (**Fig. 5B**), which locates in the cytoplasm of the host cell.

We checked whether these 4 highly immunogenic peptides/ epitopes were also revealed in other studies. It showed that S1-5 was identified by a peptide microarray study (*22*). S2-78 was identified by ReScan (*23*), and 3 peptides (S2-78, S2-96, S2-97) were predicted by COVIDep (*21*). The functional role of S1-5 specific antibodies may worth further investigation. S2-78 is adjacent to HR2, the antibody targeting this site may block the conformational change that is essential for effective virus-cell fusion (*39*). It is surprising to see the high immunogenicity of S2-96 and S2-97, since they are at the very C-terminal

end of Spike protein and theoretically localize in the cytoplasm. Further study is necessary to explore the underlying mechanism and functional roles of these peptides/ epitopes.

High homologies were observed for all these peptides among SARS-Cov-2, SARS-CoV and BtCoV-RaTG13 (**Fig. 5C**). Interestingly, S2-78 and S2-96/97 are highly homologous among all the coronaviruses, and almost identical among SARS-Cov-2, SARS-CoV and BtCoV-RaTG13. For S2-78, high homology was also observed for MERS-CoV. Same as S2-22/23, S2-78 has the potential to serve as a promising candidate for developing broad neutralization antibody and vaccine.

## Discussion

Herein, we aim to reveal IgG responses triggered by SARS-CoV-2 spike protein on a systematic level. We adopted a newly developed peptide microarray with full coverage of Spike protein (*14*), analyzed 1,051 COVID-19 sera and 528 control sera. A set of highly immunogenic peptides/epitopes were revealed, and a comprehensive IgG linear epitope landscape was constructed.

One limitation of this study is that only short peptides were involved. Though linear epitopes are nicely represented, conformational epitopes may not, for example, for RBD region, which is highly immunogenic, only 3 linear epitopes of moderate immunogenicity were identified. To overcome this limitation, one way is to synthesize longer peptides which may retain some conformational information (*31*). It is necessary to point out that for the linear epitopes that we identified, they are highly physiologically relevant, since all of them were revealed through the analysis of sera from COVID-19 patient.

Our study presents the first IgG linear epitope landscape of Spike protein, which could only be enabled by analyzing a large cohort of samples using a systematic approach, such as the peptide microarray of full coverage of Spike protein. According to the landscape, it is obvious that Spike protein is highly immunogenic, there are many epitopes on the protein, and these epitopes are not evenly distributed across Spike protein. Among the 19 immunogenic peptides (**Table S3**), some may be good candidates for developing neutralization antibodies or vaccines, some may cause ADE, if any. The rest of the peptides may have no direct biological function, but serve only as neutral immunodominant epitopes

that could not trigger neutralization activity and ADE. We believe that most of the 19 immunogenic peptides (**Table S3**) are worth further testing in animal model to identify which peptides can specifically trigger neutralization activities. When use the intact Spike protein for vaccine development, we can take a more precise way by deleting or mutating the peptides that correspond to ADE and neutral immunodominant epitopes, thus assure the efficacy while minimize the side effect. Alternatively, the peptides which elicit neutralization activity have the potential to be applied directly for neutralization antibody generation, or applied as peptide vaccine. By this way, we can avoid many challenges when working with intact Spike protein or RBD.

The linear epitope landscape and the highly immunogenic peptides identified in this study could serve as a solid base to guide the precise development of prophylactic/ therapeutic antibodies and vaccines for combating COVID-19. Because of the high homology, they may also effective for other human coronaviruses, including SARS-CoV and MERS-CoV. Since most of the highly immunogenic peptides are outside of RBD, we believe that the antibodies and vaccines based on these peptides/epitopes could be used alone or at least as a complementary to the RBD centered antibodies and vaccines.

## Materials and Methods

### *Peptide synthesis and conjugation with BSA*

The N-terminal amidated peptides were synthesized by GL Biochem, Ltd. (Shanghai, China). Each peptide was individually conjugated with BSA using Sulfo-SMCC (Thermo Fisher Scientific, MA, USA) according to the manufacture's instruction. Briefly, BSA was activated by Sulfo-SMCC in a molar ratio of 1: 30, followed by dialysis in PBS buffer. The peptide with cysteine was added in a w/w ratio of 1:1 and incubated for 2 h, followed by dialysis in PBS to remove free peptides. A few conjugates were randomly selected for examination by SDS-PAGE. For the conjugates of biotin-BSA-peptide, before conjugation, BSA was labelled with biotin by using NHS-LC-Biotin reagent (Thermo Fisher Scientific, MA, USA) with a molar ratio of 1: 5, and then activated by Sulfo-SMCC.

### *Peptide microarray fabrication*

The peptide-BSA conjugates as well as S1 protein, RBD protein and N protein of SARS-CoV-2, along with the negative (BSA) and positive controls (anti-Human IgG and IgM antibody), were printed in triplicate on PATH substrate slide (Grace Bio-Labs, Oregon, USA) to generate identical arrays in a 1 x 7 subarray format using Super Marathon printer (Arrayjet, UK). The microarrays were stored at -80 ℃ until use.

### *Patients and samples*

The study was approved by the Ethical Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China (IRB ID:TJ-C20200128) Written informed consent was obtained from all participants enrolled in this study. COVID-19 patients were hospitalized and received treatment in Tongji Hospital during the period from 17 February 2020 and 28 April 2020. Serum from each patient was collected when hospitalization at viable time points (**Table S1**). Sera of the control group from healthy donors, lung cancer patients, patients with autoimmune diseases were collected from Ruijin Hospital, Shanghai, China or Tongren Hospital, Shanghai, China. All the sera were stored at -80℃ until use.

### *Microarray-based serum analysis*

A 14-chamber rubber gasket was mounted onto each slide to create individual chambers for the 14 identical subarrays. The microarray was used for serum profiling as described previously with minor modifications(*40*). Briefly, the arrays stored at -80 ℃ were warmed to room temperature and then incubated in blocking buffer (3% BSA in 1 ×PBS buffer with 0.1% Tween 20) for 3 h. A total of 200 μL of diluted sera or antibodies was incubated with each subarray for 2 h. The sera were diluted at 1:200 for most samples and for competition experiment, free peptides were added at a concentration of 0.25 mg/mL. For the enriched antibodies, 0.1-0.5 μg antibodies were included in 200 μL incubation buffer. The arrays were washed with 1 ×PBST and bound antibodies were detected by incubating with Cy3-conjugated goat anti-human IgG and Alexa Fluor 647-conjugated donkey anti-human IgM

(Jackson ImmunoResearch, PA, USA), which were diluted for 1: 1,000 in 1×PBST. The incubation was carried out at room temperature for 1 h. The microarrays were then washed with 1×PBST and dried by centrifugation at room temperature and scanned by LuxScan 10K-A (CapitalBio Corporation, Beijing, China) with the parameters set as 95% laser power/ PMT 550 and 95% laser power/ PMT 480 for IgM and IgG, respectively. The fluorescent intensity was extracted by GenePix Pro 6.0 software (Molecular Devices, CA, USA).

***Data analysis of peptide microarray***

For each spot, signal intensity was defined as the mean_foreground subtracted by the mean_background. The signal intensities of the triplicate spots for each peptide or protein were averaged. The overall_mean_background and the overall_standard deviation (SD)_background of all the arrays probed with COVID-19 sera were calculated. Cutoff1 was defined as (the overall_mean_background + 2*overall_SD_background). According to the array data, Cutoff1 was calculated as 380.7. For the control arrays, mean_forground and SD_foreground for each peptide and protein were calculated. Cutoff2 was set as (control_mean_singal intensity + 2*control_SD_ signal intentisy). For each peptide or protein, SARS-CoV-2 specific positive response was called when the average_signal intesnty is larger than both Cutoff1 and Cutoff2. Response frequency was then defined as the number of the peptides with positive response divided by the total number of the peptides on the microarray.

***Structure analysis***

The spike protein structures (PDB ID: 6X6P and 6VYB), RBD-ACE2 structure (PDB ID: 6M0J) and antibodies-RBD complex structure (PDB ID: 7C01, 7BWJ, 7BYR, 6W41 and 6WPT) were used to analyze the structural details of the epitopes identified from the peptide microarray. The C terminal (1146-1273) structure of Spike protein was from a modeling structure, QHD43416.pdb, generated by the C-I-TASSER (https://zhanglab.ccmb.med.umich.edu/COVID-19/), and aligned to C terminal of Spike

protein (PDB ID: 6X6P). Structural analysis was processed in Pymol. The alignment and homology analysis of 7 human coronaviruses and one bat coronavirus was generated by ClustalW algorithm from EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/).

**Acknowledgments:** We thank Dr. Daniel M. Czajkowsky for critical reading and editing.

**Funding**: This work was partially supported by National Key Research and Development Program of China Grant (No. 2016YFA0500600), Science and Technology Commission of Shanghai Municipality (No. 19441911900), Interdisciplinary Program of Shanghai Jiao Tong University (No. YG2020YQ10), National Natural Science Foundation of China (No. 31970130, 31600672, 31670831, and 31370813).

**Author contributions**: S-C. T. developed the conceptual ideas and designed the study. Z-Y. S., F.W., H-Y. H., Y-D. Z., X-S. L., Z-J. Y., H-M. S. and J-X. W. collected the sera samples. X-L F., Y. L., M-L. M., Z-W. X., B. Z., H. C., C-Z. Y., J-B. X., X-N. W., Y-X. Z., D-Y. L., H-N. Z., H-W. J., H. Q., and S-J. G. performed the experiments. S-C.T., Y. L. and M-L. M. wrote the manuscript with suggestions from other authors.

**Competing interests**: The authors declare no competing interest.

**Data and materials availability**: Additional data related to this paper may be requested from the authors.

## References

1. P. Zhou, X. Lou Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, H. D. Chen, J. Chen, Y. Luo, H. Guo, R. Di Jiang, M. Q. Liu, Y. Chen, X. R. Shen, X. Wang, X. S. Zheng, K. Zhao, Q. J. Chen, F. Deng, L. L. Liu, B. Yan, F. X. Zhan, Y. Y. Wang, G. F. Xiao, Z. L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. **579**, 270–273 (2020).
2. F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, Y. L. Zhang, F. H. Dai, Y. Liu, Q. M. Wang, J. J. Zheng, L. Xu, E. C. Holmes, Y. Z. Zhang, A new coronavirus associated with human respiratory disease in China. Nature. 579, 265–269 (2020).
3. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases. 20, 533–534 (2020).
4. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. Nature Medicine. 26, 450–452 (2020).
5. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike. Science, eabb9983 (2020).
6. Y. Wu, F. Wang, C. Shen, W. Peng, D. Li, C. Zhao, Z. Li, S. Li, Y. Bi, Y. Yang, Y. Gong, H. Xiao, Z. Fan, S. Tan, G. Wu, W. Tan, X. Lu, C. Fan, Q. Wang, Y. Liu, C. Zhang, J. Qi, G. F. Gao, F. Gao, L. Liu, A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. Science. 368, 1274–1278 (2020).
7. Y. Cao, B. Su, X. Guo, W. Sun, Y. Deng, L. Bao, Q. Zhu, X. Zhang, Y. Zheng, C. Geng, X. Chai, R. He, X. Li, Q. Lv, H. Zhu, W. Deng, Y. Xu, Y. Wang, L. Qiao, Y. Tan, L. Song, G. Wang, X. Du, N. Gao, J. Liu, J. Xiao, X. Su, Z. Du, Y. Feng, C. Qin, C. Qin, R. Jin, X. S. Xie, Cell, in press.
8. M. Yuan, N. C. Wu, X. Zhu, C. C. D. Lee, R. T. Y. So, H. Lv, C. K. P. Mok, I. A. Wilson, A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science. 368, 630–633 (2020).
9. A. Z. Wec, D. Wrapp, A. S. Herbert, D. P. Maurer, D. Haslwanter, M. Sakharkar, R. K. Jangra, M. E. Dieterle, A. Lilov, D. Huang, L. V Tse, N. V Johnson, C.-L. Hsieh, N. Wang, J. H. Nett, E. Champney, I. Burnina, M. Brown, S. Lin, M. Sinclair, C. Johnson, S. Pudi, R. Bortz, A. S. Wirchnianski, E. Laudermilch, C. Florez, J. M. Fels, C. M. O'Brien, B. S. Graham, D. Nemazee, D. R. Burton, R. S. Baric, J. E. Voss, K. Chandran, J. M. Dye, J. S. McLellan, L. M. Walker, Broad neutralization of SARS-related viruses by human monoclonal antibodies. Science, eabc7424 (2020).
10. A. Baum, B. O. Fulton, E. Wloga, R. Copin, K. E. Pascal, V. Russo, S. Giordano, K. Lanza, N. Negron, M. Ni, Y. Wei, G. S. Atwal, A. J. Murphy, N. Stahl, G. D. Yancopoulos, C. A. Kyratsous, Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science, eabd0831 (2020).
11. J. Hansen, A. Baum, K. E. Pascal, V. Russo, S. Giordano, E. Wloga, B. O. Fulton, Y. Yan, K. Koon, K. Patel, K. M. Chung, A. Hermann, E. Ullman, J. Cruz, A.

Rafique, T. Huang, J. Fairhurst, C. Libertiny, M. Malbec, W.-Y. Lee, R. Welsh, G. Farr, S. Pennington, D. Deshpande, J. Cheng, A. Watty, P. Bouffard, R. Babb, N. Levenkova, C. Chen, B. Zhang, A. Romero Hernandez, K. Saotome, Y. Zhou, M. Franklin, S. Sivapalasingam, D. C. Lye, S. Weston, J. Logue, R. Haupt, M. Frieman, G. Chen, W. Olson, A. J. Murphy, N. Stahl, G. D. Yancopoulos, C. A. Kyratsous, Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. Science, eabd0827 (2020).

12. T. F. Rogers, F. Zhao, D. Huang, N. Beutler, A. Burns, W.-T. He, O. Limbo, C. Smith, G. Song, J. Woehl, L. Yang, R. K. Abbott, S. Callaghan, E. Garcia, J. Hurtado, M. Parren, L. Peng, S. Ramirez, J. Ricketts, M. J. Ricciardi, S. A. Rawlings, N. C. Wu, M. Yuan, D. M. Smith, D. Nemazee, J. R. Teijaro, J. E. Voss, I. A. Wilson, R. Andrabi, B. Briney, E. Landais, D. Sok, J. G. Jardine, D. R. Burton, Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. Science, eabc7520 (2020).
13. P. J. M. Brouwer, T. G. Caniels, K. van der Straten, J. L. Snitselaar, Y. Aldon, S. Bangaru, J. L. Torres, N. M. A. Okba, M. Claireaux, G. Kerster, A. E. H. Bentlage, M. M. van Haaren, D. Guerra, J. A. Burger, E. E. Schermer, K. D. Verheul, N. van der Velde, A. van der Kooi, J. van Schooten, M. J. van Breemen, T. P. L. Bijl, K. Sliepen, A. Aartse, R. Derking, I. Bontjer, N. A. Kootstra, W. J. Wiersinga, G. Vidarsson, B. L. Haagmans, A. B. Ward, G. J. de Bree, R. W. Sanders, M. J. van Gils, Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. Science, eabc5902 (2020).
14. Y. Li, D. Lai, H. Zhang, H. Jiang, X. Tian, M. Ma, H. Qi, Q. Meng, S. Guo, Y. Wu, W. Wang, X. Yang, D. Shi, J. Dai, T. Ying, J. Zhou, S. Tao, Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. medRxiv (2020), doi:10.1101/2020.06.07.20125096.
15. C. M. Poh, G. Carissimo, B. Wang, S. N. Amrun, Y.-P. Lee, R. S.-L. Chee, N. K.-W. Yeo, W.-H. Lee, Y.-S. Leo, M. I.-C. Chen, S.-Y. Tan, L. Y. A. Cnhai, S. Kalimuddin, S.-Y. Thien, B. E. Young, D. C. Lye, C.-I. Wang, L. Renia, L. F. P. Ng, Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. Nature Communications. 11, 2806 (2020).
16. X. Chi, R. Yan, J. Zhang, G. Zhang, Y. Zhang, M. Hao, Z. Zhang, P. Fan, Y. Dong, Y. Yang, Z. Chen, Y. Guo, J. Zhang, Y. Li, X. Song, Y. Chen, L. Xia, L. Fu, L. Hou, J. Xu, C. Yu, J. Li, Q. Zhou, W. Chen, Science, in press.
17. L. Liu, Q. Wei, Q. Lin, J. Fang, H. Wang, H. Kwok, H. Tang, K. Nishiura, J. Peng, Z. Tan, T. Wu, K. W. Cheung, K. H. Chan, X. Alvarez, C. Qin, A. Lackner, S. Perlman, K. Y. Yuen, Z. Chen, Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. JCI insight. 4, e123158 (2019).
18. N. Vabret, G. J. Britton, C. Gruber, S. Hegde, J. Kim, M. Kuksin, R. Levantovsky, L. Malle, A. Moreira, M. D. Park, L. Pia, E. Risson, M. Saffern, B. Salomé, M. Esai Selvan, M. P. Spindler, J. Tan, V. van der Heide, J. K. Gregory, K. Alexandropoulos, N. Bhardwaj, B. D. Brown, B. Greenbaum, Z. H. Gümüş, D. Homann, A. Horowitz, A. O. Kamphorst, M. A. Curotto de Lafaille, S. Mehandru, M. Merad, R. M. Samstein, M. Agrawal, M. Aleynick, M. Belabed, M. Brown, M. Casanova-Acebes, J. Catalan, M. Centa, A. Charap, A. Chan, S. T. Chen, J.

Chung, C. C. Bozkus, E. Cody, F. Cossarini, E. Dalla, N. Fernandez, J. Grout, D. F. Ruan, P. Hamon, E. Humblin, D. Jha, J. Kodysh, A. Leader, M. Lin, K. Lindblad, D. Lozano-Ojalvo, G. Lubitz, A. Magen, Z. Mahmood, G. Martinez-Delgado, J. Mateus-Tique, E. Meritt, C. Moon, J. Noel, T. O'Donnell, M. Ota, T. Plitt, V. Pothula, J. Redes, I. Reyes Torres, M. Roberto, A. R. Sanchez-Paulete, J. Shang, A. S. Schanoski, M. Suprun, M. Tran, N. Vaninov, C. M. Wilk, J. Aguirre-Ghiso, D. Bogunovic, J. Cho, J. Faith, E. Grasset, P. Heeger, E. Kenigsberg, F. Krammer, U. Laserson, Immunology of COVID-19: Current State of the Science. Immunity. 52, 910–941 (2020).

19. M. Sikora, S. von Bülow, F. E. C. Blanc, M. Gecht, R. Covino, G. Hummer, Map of SARS-CoV-2 spike epitopes not shielded by glycans. bioRxiv (2020), doi:10.1101/2020.07.03.186825.
20. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology. 157, 105–132 (1982).
21. S. F. Ahmed, A. A. Quadeer, M. R. McKay, COVIDep: A web-based platform for real-time reporting of vaccine target recommendations for SARS-CoV-2. Nature Protocols. 15, 2141–2142 (2020).
22. H. Wang, X. Hou, X. Wu, T. Liang, X. Zhang, D. Wang, F. Teng, J. Dai, H. Duan, S. Guo, Y. Li, X. Yu, SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution. bioRxiv (2020), doi:10.1101/2020.03.26.994756.
23. C. R. Zamecnik, J. V Rajan, K. A. Yamauchi, S. A. Mann, G. M. Sowa, K. C. Zorn, B. D. Alvarenga, M. Stone, P. J. Norris, W. Gu, C. Y. Chiu, J. L. Derisi, M. R. Wilson, ReScan, a Multiplex Diagnostic Pipeline, Pans Human Sera for SARS-CoV-2 Antigens. medRxiv (2020), doi:10.1101/2020.05.11.20092528.
24. G. J. Xu, T. Kula, Q. Xu, M. Z. Li, S. D. Vernon, T. Ndung'u, K. Ruxrungtham, J. Sanchez, C. Brander, R. T. Chung, K. C. O'Connor, B. Walker, H. B. Larman, S. J. Elledge, Comprehensive serological profiling of human populations using a synthetic human virome. Science. 384, aaa0698 (2015).
25. Q. X. Long, B. Z. Liu, H. J. Deng, G. C. Wu, K. Deng, Y. K. Chen, P. Liao, J. F. Qiu, Y. Lin, X. F. Cai, D. Q. Wang, Y. Hu, J. H. Ren, N. Tang, Y. Y. Xu, L. H. Yu, Z. Mo, F. Gong, X. L. Zhang, W. G. Tian, L. Hu, X. X. Zhang, J. L. Xiang, H. X. Du, H. W. Liu, C. H. Lang, X. H. Luo, S. B. Wu, X. P. Cui, Z. Zhou, M. M. Zhu, J. Wang, C. J. Xue, X. F. Li, L. Wang, Z. J. Li, K. Wang, C. C. Niu, Q. J. Yang, X. J. Tang, Y. Zhang, X. M. Liu, J. J. Li, D. C. Zhang, F. Zhang, P. Liu, J. Yuan, Q. Li, J. L. Hu, J. Chen, A. L. Huang, Antibody responses to SARS-CoV-2 in patients with COVID-19. Nature Medicine. 26, 845–848 (2020).
26. H. Jiang, Y. Li, H. Zhang, W. Wang, D. Men, X. Yang, H. Qi, J. Zhou, S. Tao, Global profiling of SARS-CoV-2 specific IgG/ IgM responses of convalescents using a proteome microarray. medRxiv (2020), doi:10.1101/2020.03.20.20039495.
27. N. G. Herrera, N. C. Morano, A. Celikgil, G. I. Georgiev, R. J. Malonis, J. H. Lee, K. Tong, O. Vergnolle, A. B. Massimi, L. Y. Yen, A. J. Noble, M. Kopylov, J. B. Bonanno, S. C. Garrett-Thomson, D. B. Hayes, M. Brenowitz, S. J. Garforth, E. T. Eng, J. R. Lai, S. C. Almo, Characterization of the SARS-CoV-2 S Protein: Biophysical, Biochemical, Structural, and Antigenic Analysis. bioRxiv (2020), doi:10.1101/2020.06.14.150607.

28. E. A. Emini, J. V Hughes, D. S. Perlow, J. Boger, Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of virology. 55, 836–839 (1985).
29. L. Premkumar, B. Segovia-Chumbez, R. Jadi, D. R. Martinez, R. Raut, A. J. Markmann, C. Cornaby, L. Bartelt, S. Weiss, Y. Park, C. E. Edwards, E. Weimer, E. M. Scherer, N. Rouphael, S. Edupuganti, D. Weiskopf, L. V Tse, Y. J. Hou, D. Margolis, A. Sette, M. H. Collins, J. Schmitz, R. S. Baric, A. M. de Silva, The receptor-binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. Science Immunology. 5, eabc8413 (2020).
30. J. Wan, S. Xing, L. Ding, Y. Wang, D. Zhu, B. Rong, S. Wang, K. Chen, C. He, S. Yuan, C. Qiu, C. Zhao, X. Zhang, X. Wang, Y. Lu, J. Xu, F. Lan, Human IgG cell neutralizing monoclonal antibodies block SARS- CoV-2 infection. bioRxiv (2020), doi:https://doi.org/10.1101/2020.05.19.104117.
31. B. Zhang, Y. Hu, L. Chen, T. Yau, Y. Tong, J. Hu, J. Cai, K.-H. Chan, Y. Dou, J. Deng, X. Wang, I. F.-N. Hung, K. K.-W. To, K. Y. Yuen, J.-D. Huang, Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. Cell Research (2020), doi:10.1038/s41422-020-0366-x.
32. R. Shi, C. Shan, X. Duan, Z. Chen, P. Liu, J. Song, T. Song, X. Bi, C. Han, L. Wu, G. Gao, X. Hu, Y. Zhang, Z. Tong, W. Huang, W. J. Liu, G. Wu, B. Zhang, L. Wang, J. Qi, H. Feng, F. Wang, Q. Wang, G. F. Gao, Z. Yuan, J. Yan, A human neutralizing antibody targets the receptor binding site of SARS-CoV-2. Nature (2020), doi:10.1038/s41586-020-2381-y.
33. B. Ju, Q. Zhang, J. Ge, R. Wang, J. Sun, X. Ge, J. Yu, S. Shan, B. Zhou, S. Song, X. Tang, J. Yu, J. Lan, J. Yuan, H. Wang, J. Zhao, S. Zhang, Y. Wang, X. Shi, L. Liu, J. Zhao, X. Wang, Z. Zhang, L. Zhang, Human neutralizing antibodies elicited by SARS-CoV-2 infection. Nature (2020), doi:10.1038/s41586-020-2380-z.
34. D. Pinto, Y.-J. Park, M. Beltramello, A. C. Walls, M. A. Tortorici, S. Bianchi, S. Jaconi, K. Culap, F. Zatta, A. De Marco, A. Peter, B. Guarino, R. Spreafico, E. Cameroni, J. B. Case, R. E. Chen, C. Havenar-Daughton, G. Snell, A. Telenti, H. W. Virgin, A. Lanzavecchia, M. S. Diamond, K. Fink, D. Veesler, D. Corti, Structural and functional analysis of a potent sarbecovirus neutralizing antibody. bioRxiv (2020), doi:10.1101/2020.04.07.023903.
35. S. Jiang, C. Hillyer, L. Du, Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses. Trends in immunology. 41, 355–359 (2020).
36. B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, K. M. Hastie, M. D. Parker, D. G. Partridge, C. M. Evans, T. M. Freeman, T. I. de Silva, C. McDanal, L. G. Perez, H. Tang, A. Moon-Walker, S. P. Whelan, C. C. LaBranche, E. O. Saphire, D. C. Montefiori, A. Angyal, R. L. Brown, L. Carrilero, L. R. Green, D. C. Groves, K. J. Johnson, A. J. Keeley, B. B. Lindsey, P. J. Parsons, M. Raza, S. Rowland-Jones, N. Smith, R. M. Tucker, D. Wang, M. D. Wyles, Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell (2020), doi:https://doi.org/10.1016/j.cell.2020.06.043.
37. L. Zhang, C. B. Jackson, H. Mou, A. Ojha, E. S. Rangarajan, T. Izard, M. Farzan, H. Choe, The D614G mutation in the SARS-CoV-2 spike protein reduces S1

shedding and increases infectivity. bioRxiv (2020), doi:10.1101/2020.06.12.148726.

38. Y. Cai, D. Yin, S. Ling, X. Tian, Y. Li, Z. Xu, H. Jiang, X. Zhang, X. Wang, Y. Shi, Y. Zhang, L. Da, S. Tao, Q. Wang, J. Xu, T. Ying, J. Hong, A single dose SARS-CoV-2 simulating particle vaccine induces potent neutralizing activities. bioRxiv (2020), doi:10.1101/2020.05.14.093054.
39. S. Liu, G. Xiao, Y. Chen, Y. He, J. Niu, C. R. Escalante, H. Xiong, J. Farmar, A. K. Debnath, P. Tien, S. Jiang, Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: Implications for virus fusogenic mechanism and identification of fusion inhibitors. Lancet. 363, 938–947 (2004).
40. Y. Li, C.-Q. Li, S.-J. Guo, W. Guo, H.-W. Jiang, H.-C. Li, S.-C. Tao, Longitudinal serum autoantibody repertoire profiling identifies surgery-associated biomarkers in lung adenocarcinoma. EBioMedicine. 53, 102674 (2020).

## Figures:

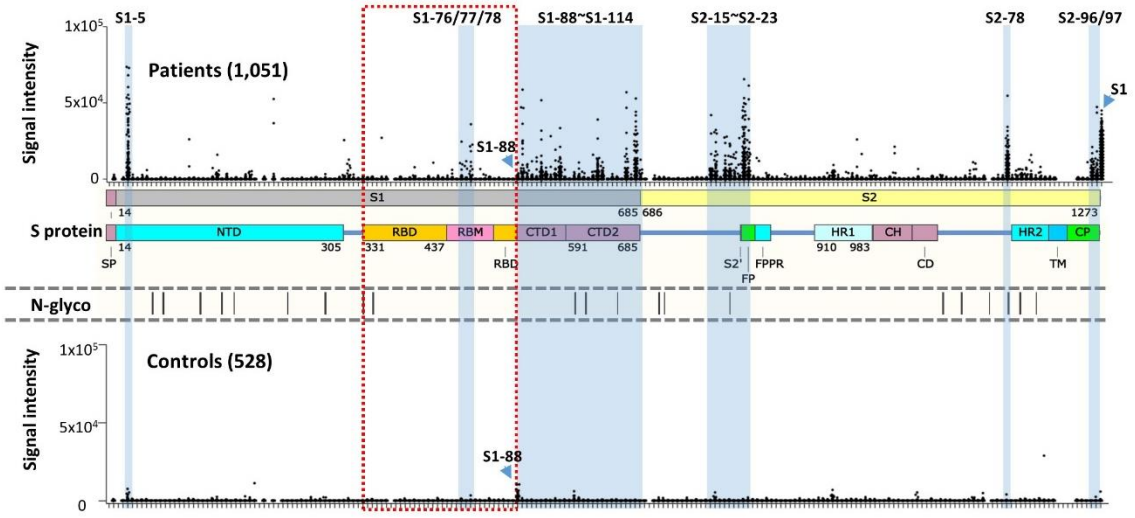


**Figure 1. The IgG linear epitope landscape of SARS-CoV-2 Spike protein.** The signal intensities of 1,051 COVID-19 sera against 197 peptides were obtained by using the peptide microarray. The peptides are listed in X-axis and aligned to the corresponding locations on Spike protein. As a control, the signal intensities of S1 protein were also presented. The missing spots are peptides either could not be synthesized or failed for BSA conjugation (see **Table S2** for details). A cohort of 528 controls sera were also analyzed on the microarray. In addition, the known N-glycosylation sites (N-glyco) were aligned with Spike protein. The peptides or regions with significant binding were marked blue. Peptide S1-88 was specifically labeled because significant bindings were also observed for the controls. SP: signaling peptide; NTD: N-terminal domain; RBD: receptor biding domain; RBM: receptor binding motif; CTD1: C-terminal domain 1; CTD2: C-terminal domain 2; S2': protease cleavage site; FP: fusion peptide; FPPR: fusion peptide proximal region; HR1: heptad repeat 1; HR2: heptad repeat 2; CH: center helix; CD: connector domain; TM: trans-membrane; CP: cytoplasmic.

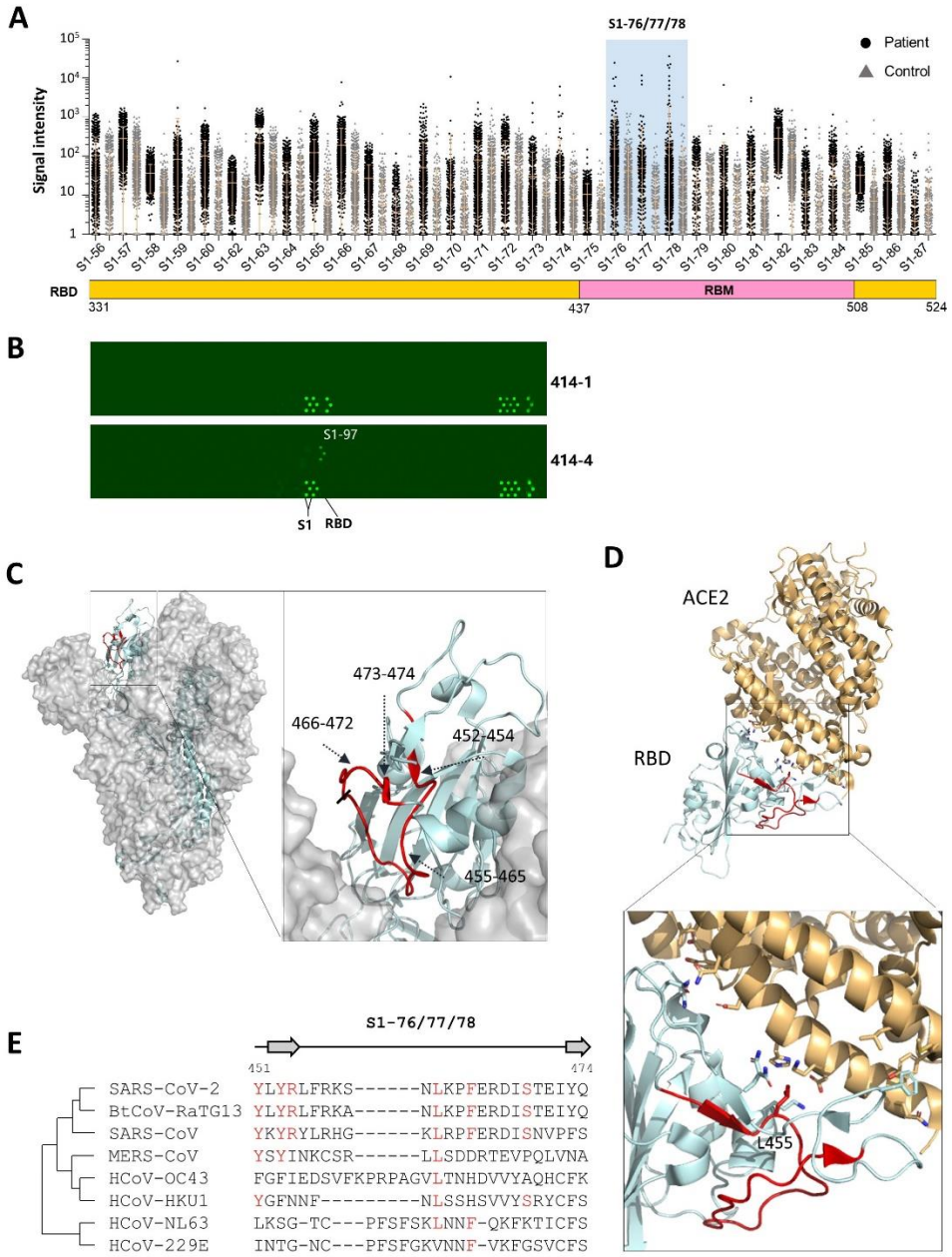


**Figure 2. RBD is lack of highly immunogenic linear epitope. A.** RBD region of the linear epitope landscape; **B**. The peptide microarray results of two Spike protein specific monoclonal human antibodies (from Active motif Co. Ltd.), one (414-1) is specific for RBD and the other (414-4) is not. **C**. The detailed structures of the significant peptides (S1-76/77/78, aa451-474, red) on RBD of the closed state Spike protein trimer, the side view (PDB ID: 6X6P); **D**. The locations of the significant peptides (S1-76/77/78, aa451-474, red) on the co-crystal structure of RBD and ACE2 (PDB ID: 6M0J). **E**. The homology analysis of the significant peptides among the 7 known human coronaviruses and the Bat coronavirus BtCoV-RaTG13, which is highly homologous to SARS-CoV-2, the amino acids with consistencies >=50% among the 8 coronaviruses are marked red, the loop and β-strand region are shown as line and arrow above the sequences, respectively.

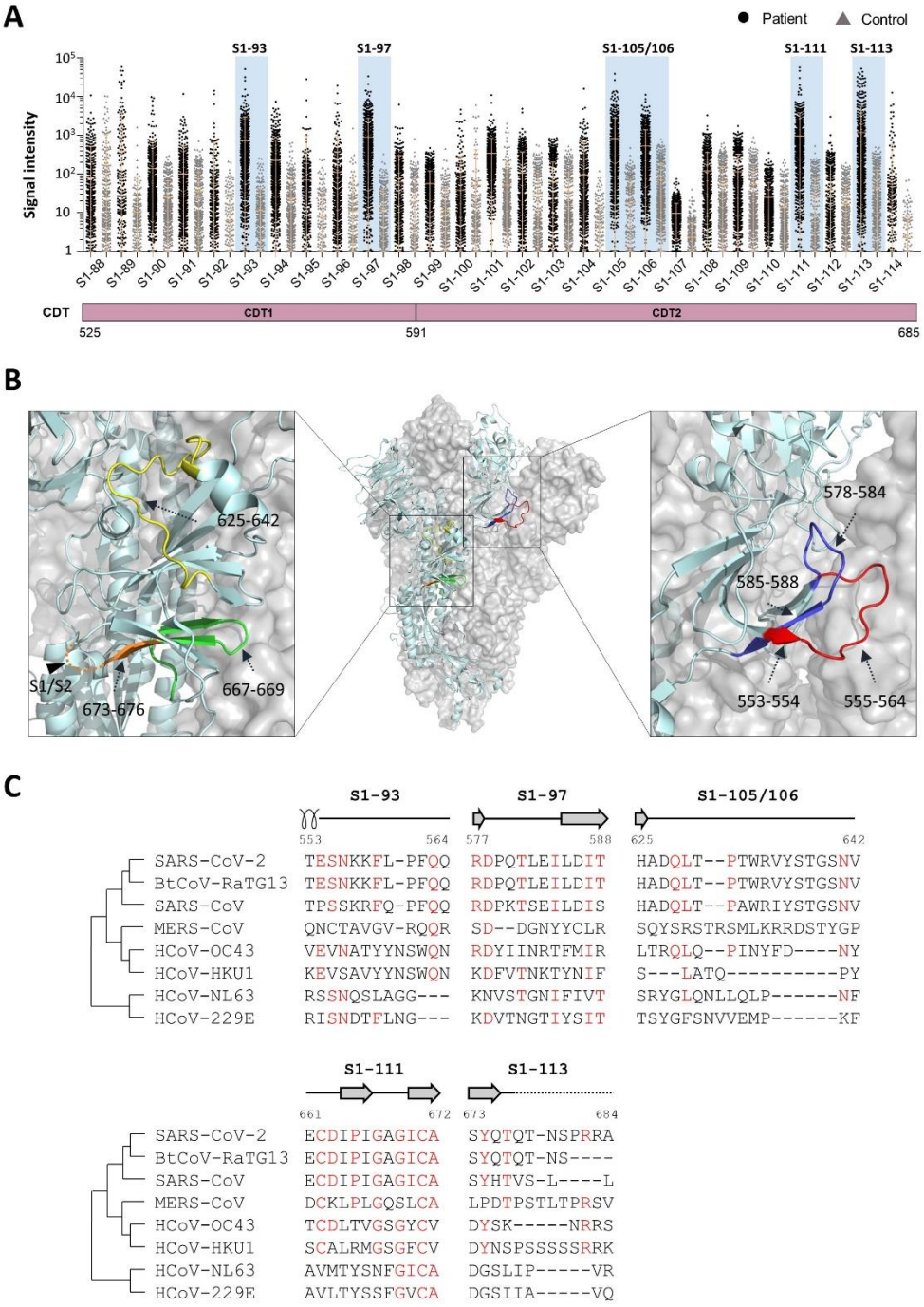


**Figure 3. CTD is rich of significant linear epitopes. A.** CTD region of the linear epitope landscape; **B**. The locations of the significant peptides locate on CTD (PDB ID: 6X6P). Specifically, S1-93, aa553-564, red; S1-97, aa577-588, blue; S1-105/106, aa625-642, yellow; S1-111, aa661-672, green; S1-113, S1-113, 673-684, orange; **C**. The homology analysis of the significant peptides among the 7 known human coronaviruses and the Bat coronavirus BtCoV-RaTG13. The amino acids with consistencies >=50% among the 8 coronaviruses are marked red, the loop, α-helix and β-strand region are shown as line, coil and arrow above sequences, respectively. Unobserved structure is shown as dotted line.

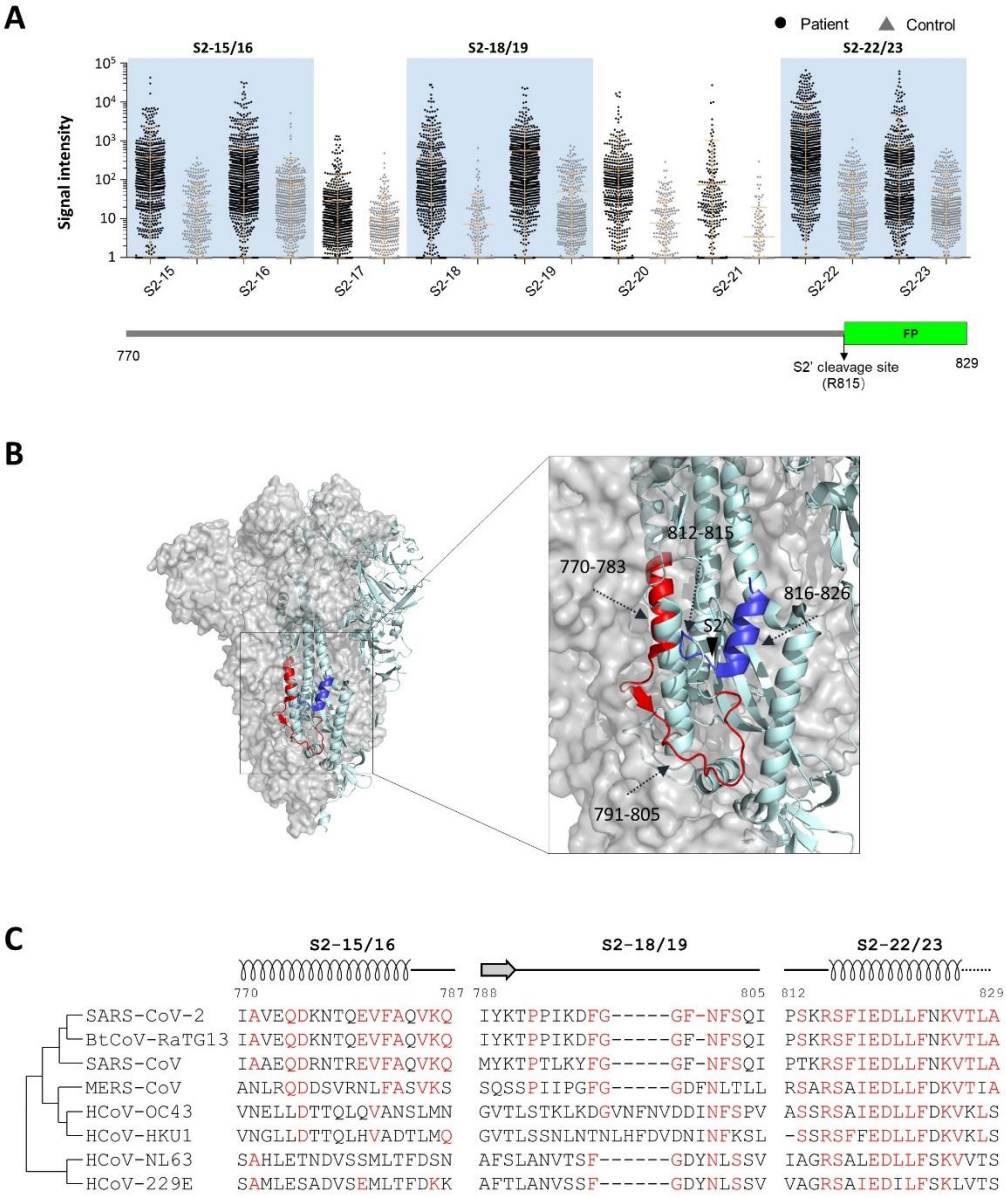


**Figure 4. The 2nd hot area of highly immunogenic linear epitopes: S2'cleavage site and FP. A.** The S2'cleavage site and FP of the linear epitope landscape; **B**. The significant peptides locate on this region. S2-15/16, aa770-787, red, coil; S2-18/19, aa788-805, red, loop; S2-22/23, aa812-829, blue, coil; **C**. The homology analysis of the significant peptides among the 7 known human coronaviruses and the Bat coronavirus BtCoV-RaTG13. The amino acids with consistencies >=50% among the 8 coronaviruses are marked red. The loop, α-helix and β-strand region are shown as line, coil and arrow above sequences, respectively. Unobserved structure is shown as dotted line.

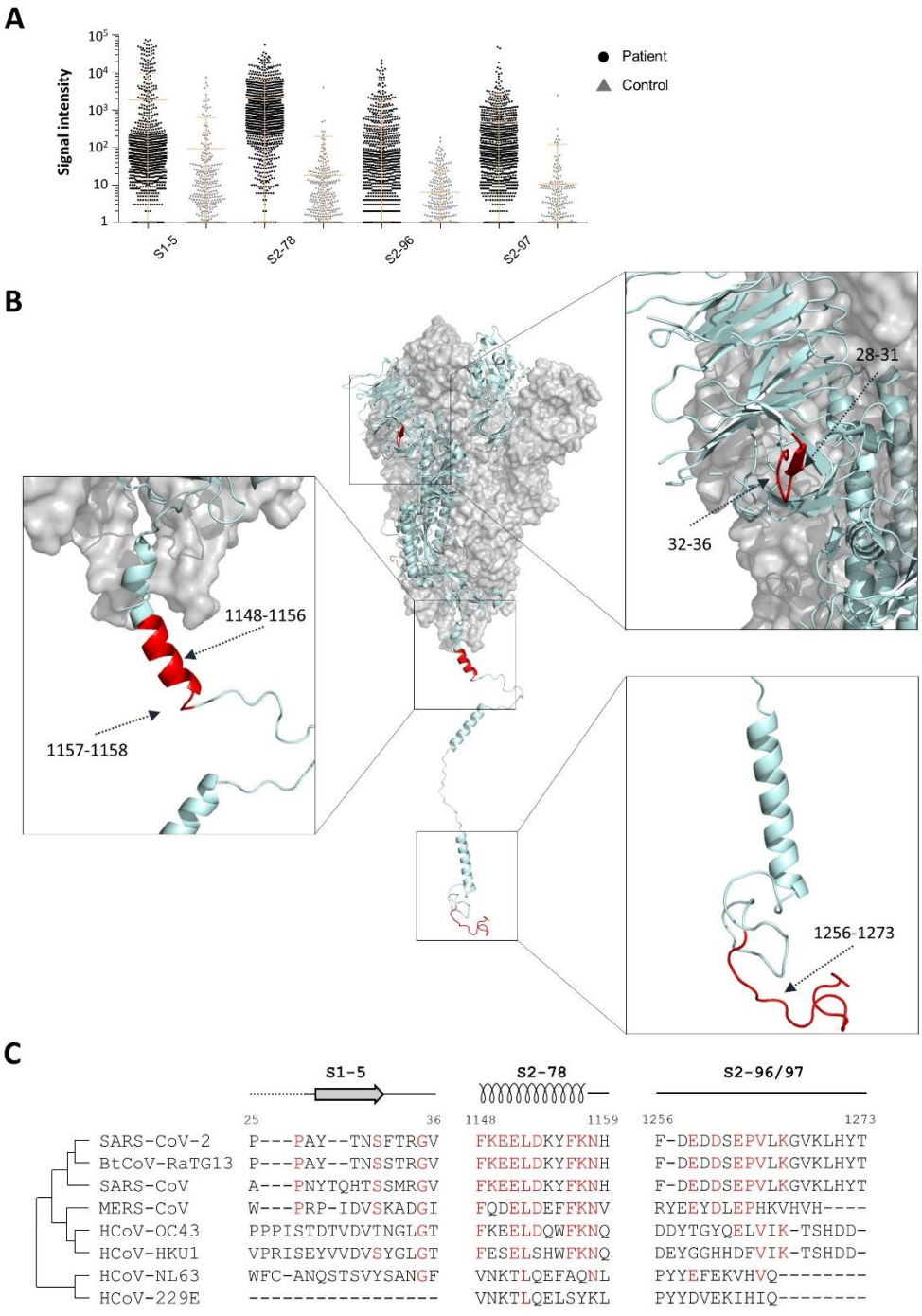


**Figure 5. Other highly immunogenic linear epitopes. A**. Other 5 significant peptides which are not belong the two "hot regions" (see **Figure 3** and **Figure 4**); **B**. The significant peptides locate on Spike protein. S1-5, aa25-36, red; S2-78, aa1148-1159, red; S2-96/97, aa1256-1273, red; **C**. The homology analysis of the significant peptides among the 7 known human coronaviruses and the Bat coronavirus BtCoV-RaTG13. The amino acids with consistencies >=50% among the 8 coronaviruses are marked red. The loop, α-helix and β-strand region are shown as line, coil and arrow above sequences, respectively. Unobserved structure is shown as dotted line.

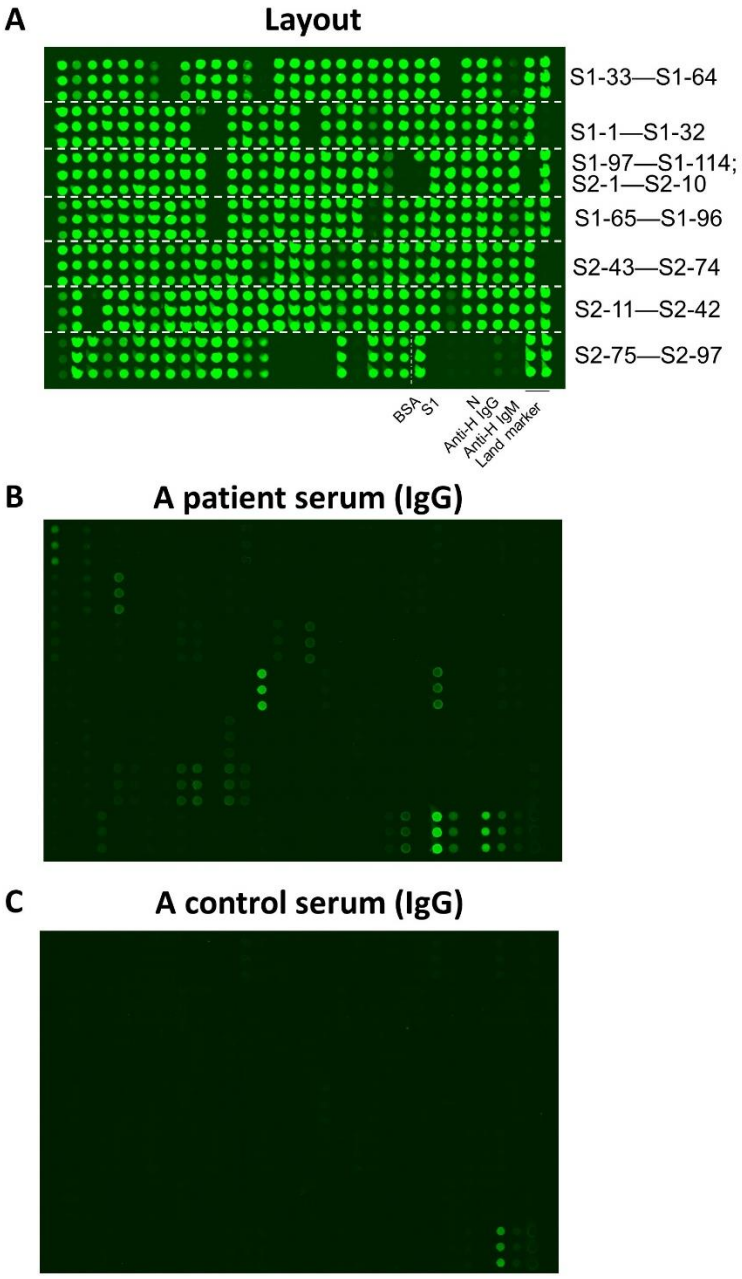


**Figure S1. The peptide microarray. A**. The layout of the peptide microarray that was used in this study (Please see **Table S2** for the details of the peptides on the array); **B**. An example array probed with COVID-19 patient serum; **C**. An example array probed with a control serum.

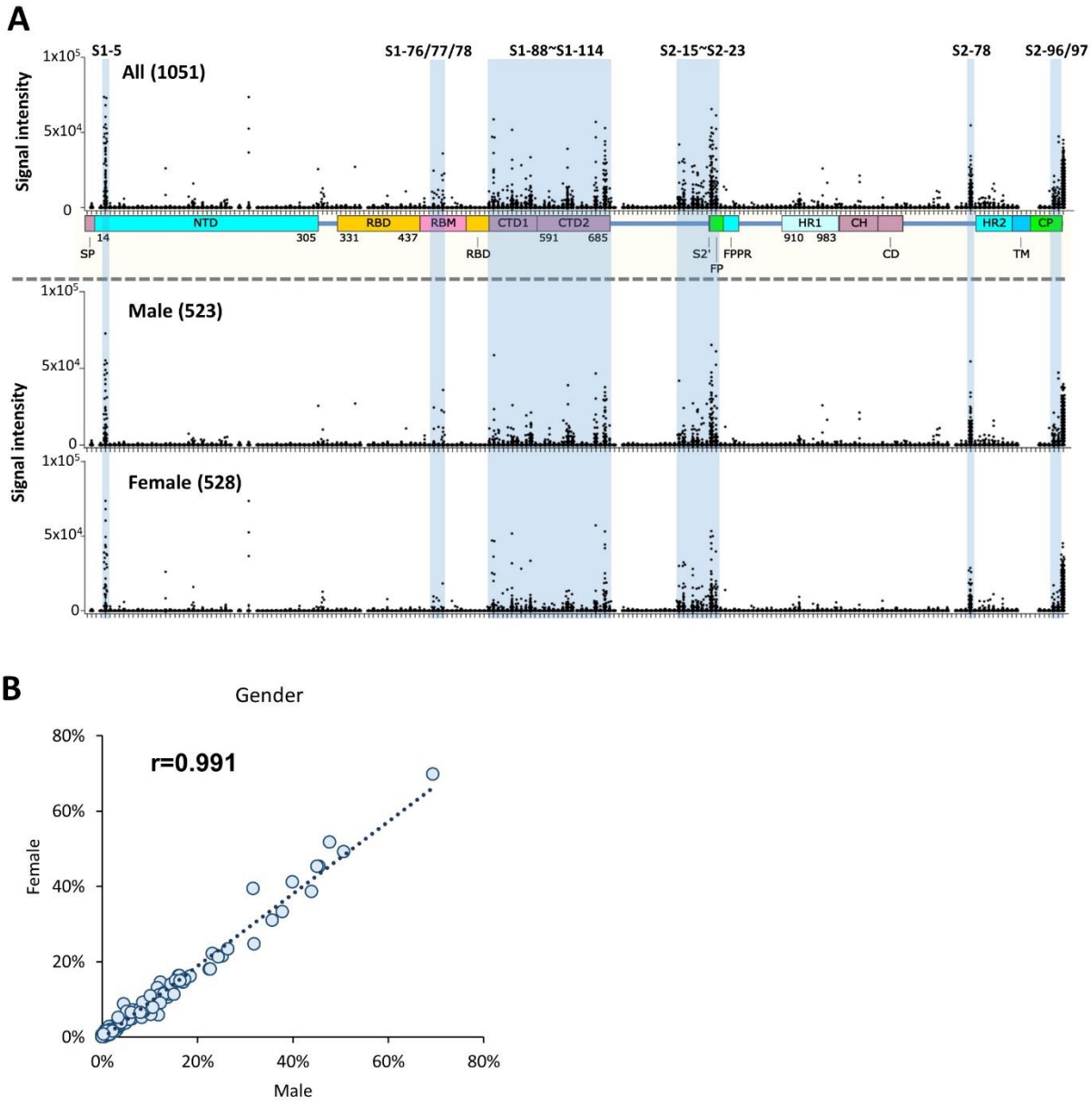


**Figure S2. The IgG linear epitope landscape of SARS-CoV-2 Spike protein: male vs. female. A.** The linear epitope landscape was divided into two sub-landscapes according to gender, *i.e.*, 523 male vs. 528 female; **B.** The correlation between male and female.

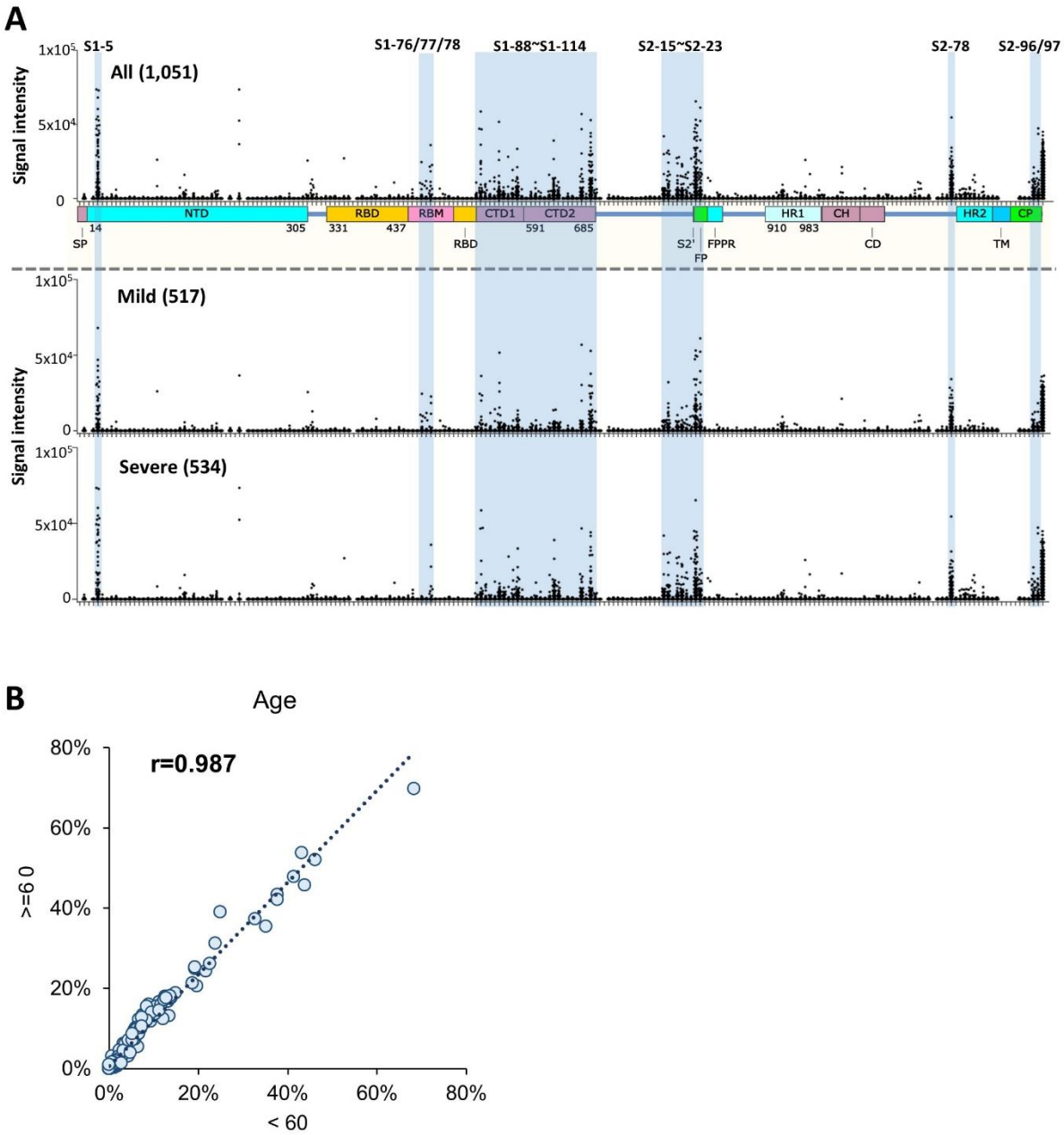


**Figure S3. The IgG linear epitope landscape of SARS-CoV-2 Spike protein: mild vs. severe. A.** The linear epitope landscape was divided into two sub-landscapes according to severity, *i.e.*, 513 mild vs 534 severe (including severe and critical cases); **B.** The correlation between mild and severe.

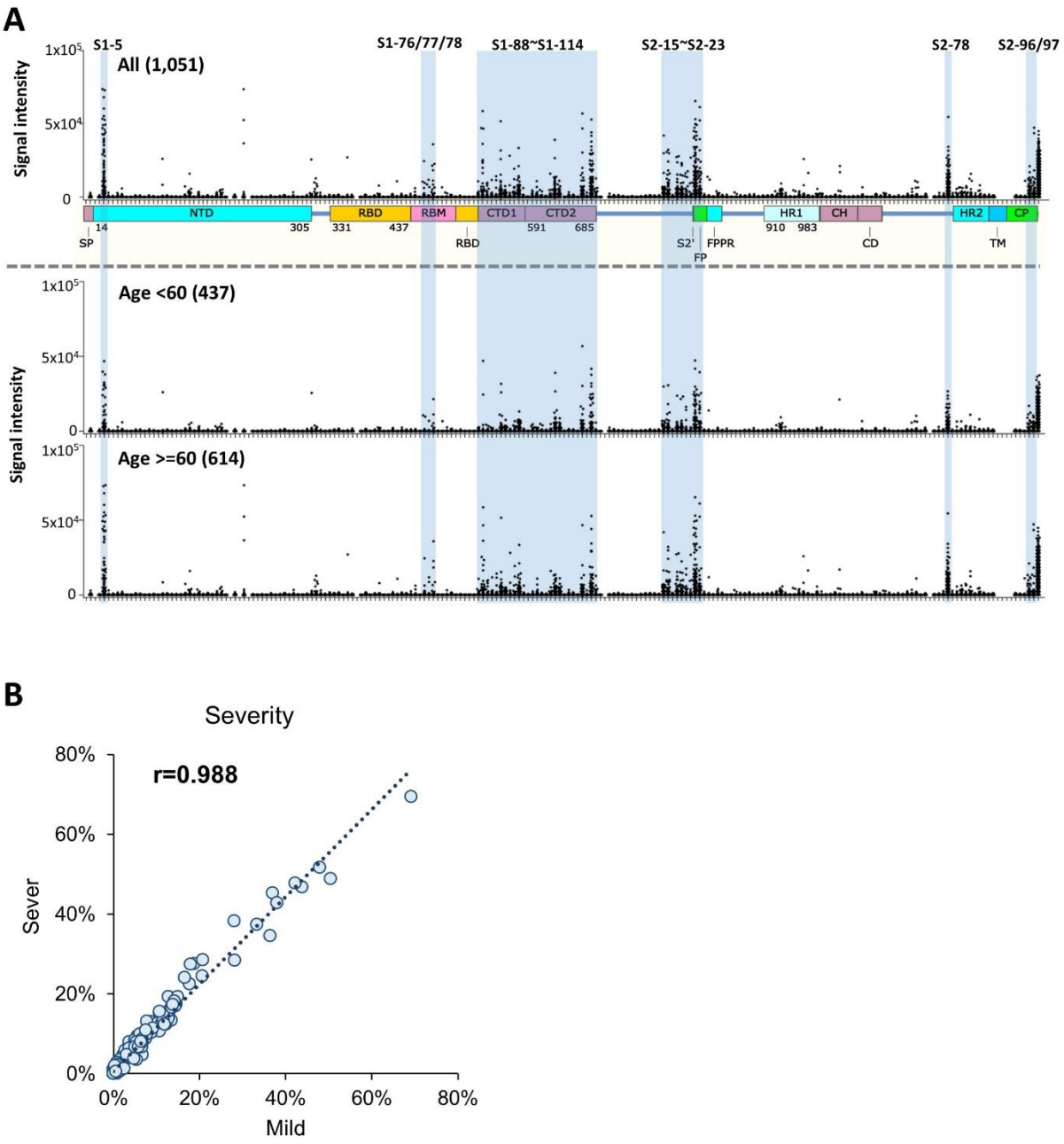


**Figure S4. The IgG linear epitope landscape of SARS-CoV-2 Spike protein: age < 60 vs. >= 60. A.** The linear epitope landscape was separated to two sub-landscapes according to age, *i.e.*, 437 age < 60 vs. 614 age >= 60; **B.** The correlation between age >=60 and age < 60.

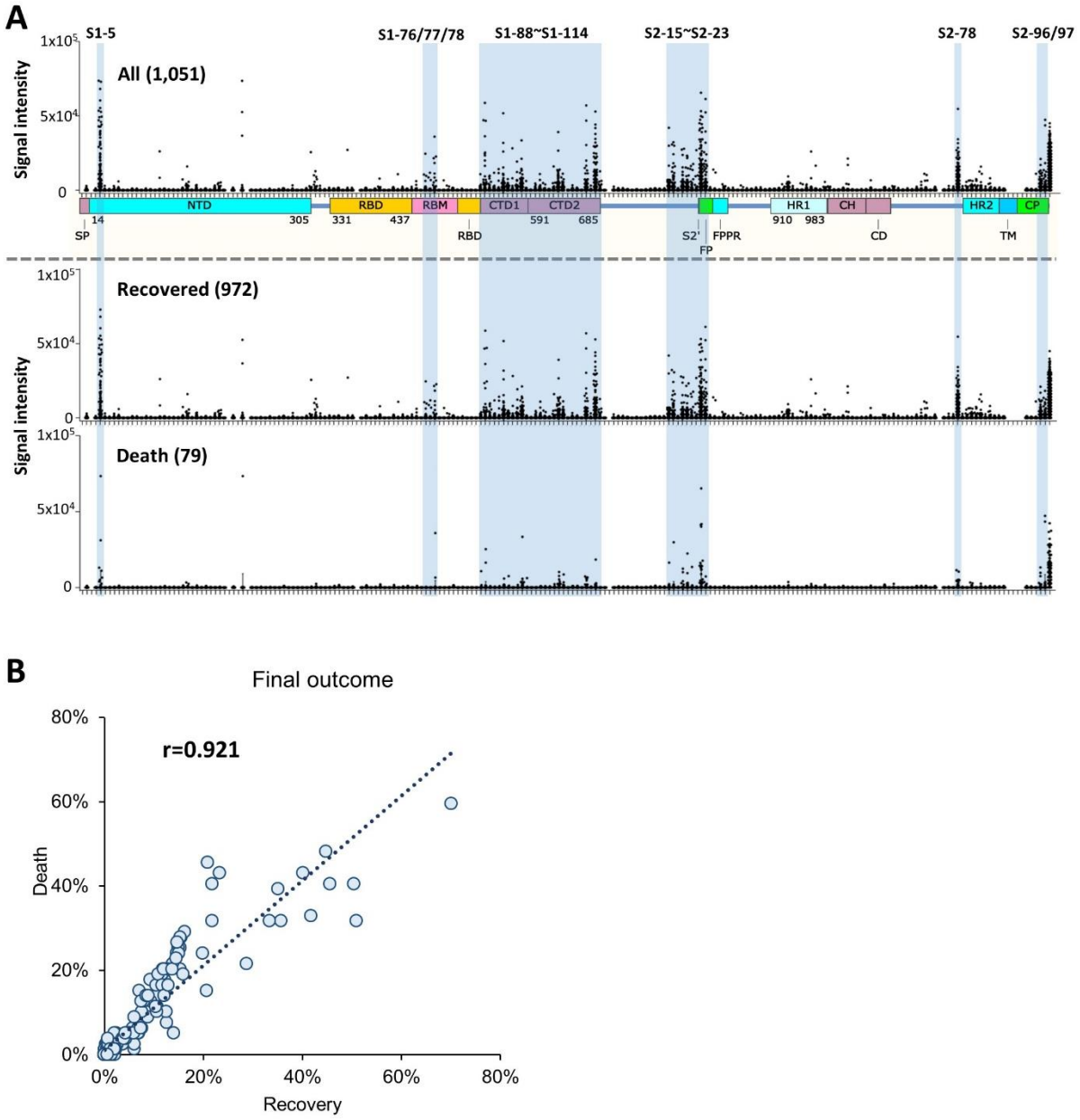


**Figure S5. The IgG linear epitope landscape of SARS-CoV-2 Spike protein: recovered vs. death. A.** The linear epitope landscape was divided into two sub-landscapes according to the final outcome, *i.e.*, recovered (972) vs. death (79); **B.** The correlation between recovered and death.

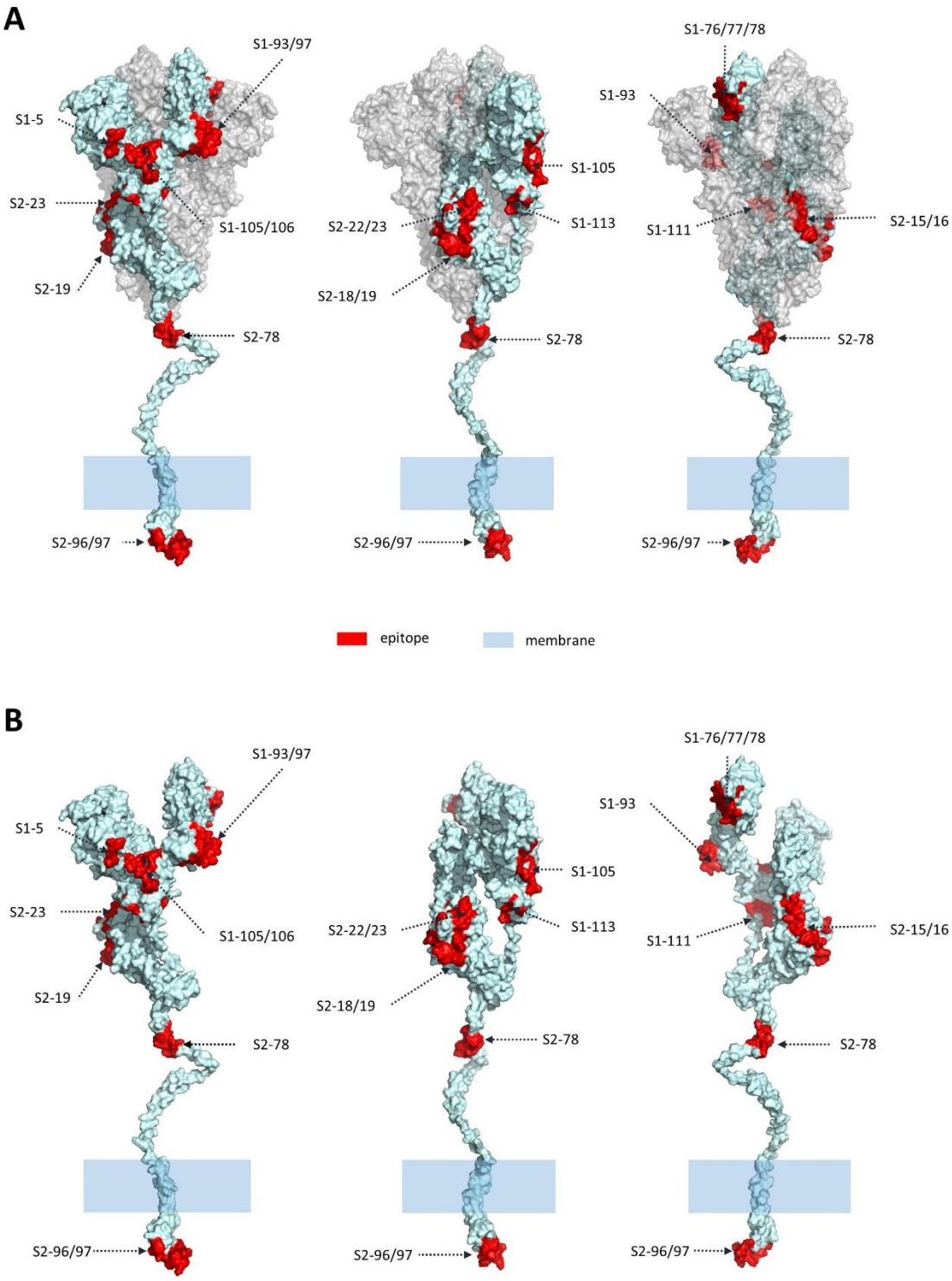


**Figure S6. The distribution of the highly immunogenic peptides on Spike protein.** A 3D structure of spike protein is adopted (PDB ID: 6X6P). The part from S2-78 to the end of C-terminal was modeled by C-I-TASSER. The 19 significant peptides are red marked on the 3D structure of Spike protein for both trimer (**A**) and monomer (**B**).

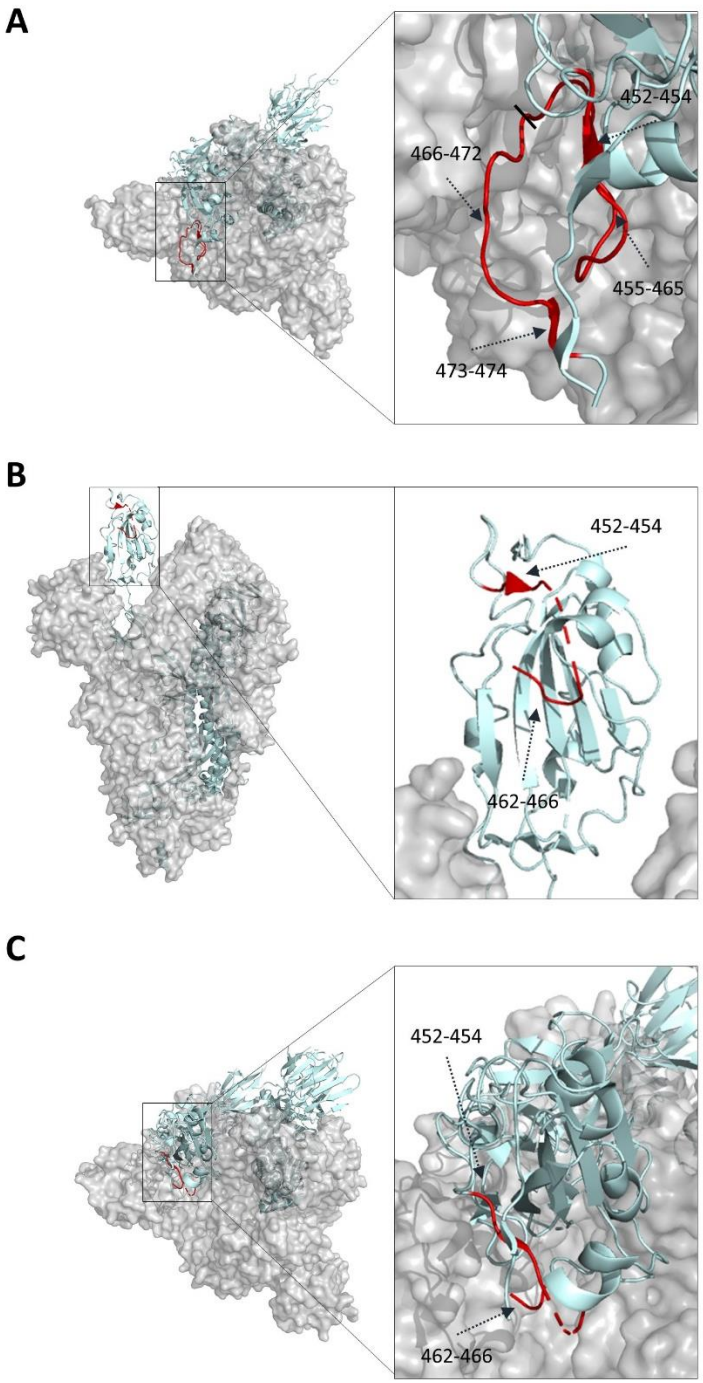


**Figure S7. The location of S1-76/77/78 on RBD. A.** The top view of closed state Spike protein trimer (PDB: 6X6P); **B.** The side view of open state Spike protein trimer (PDB ID: 6VYB); **C.** The top view of open state Spike protein trimer (PDB ID: 6VYB). The significant peptides (S1-76/77/78, aa451-474) were marked as red.

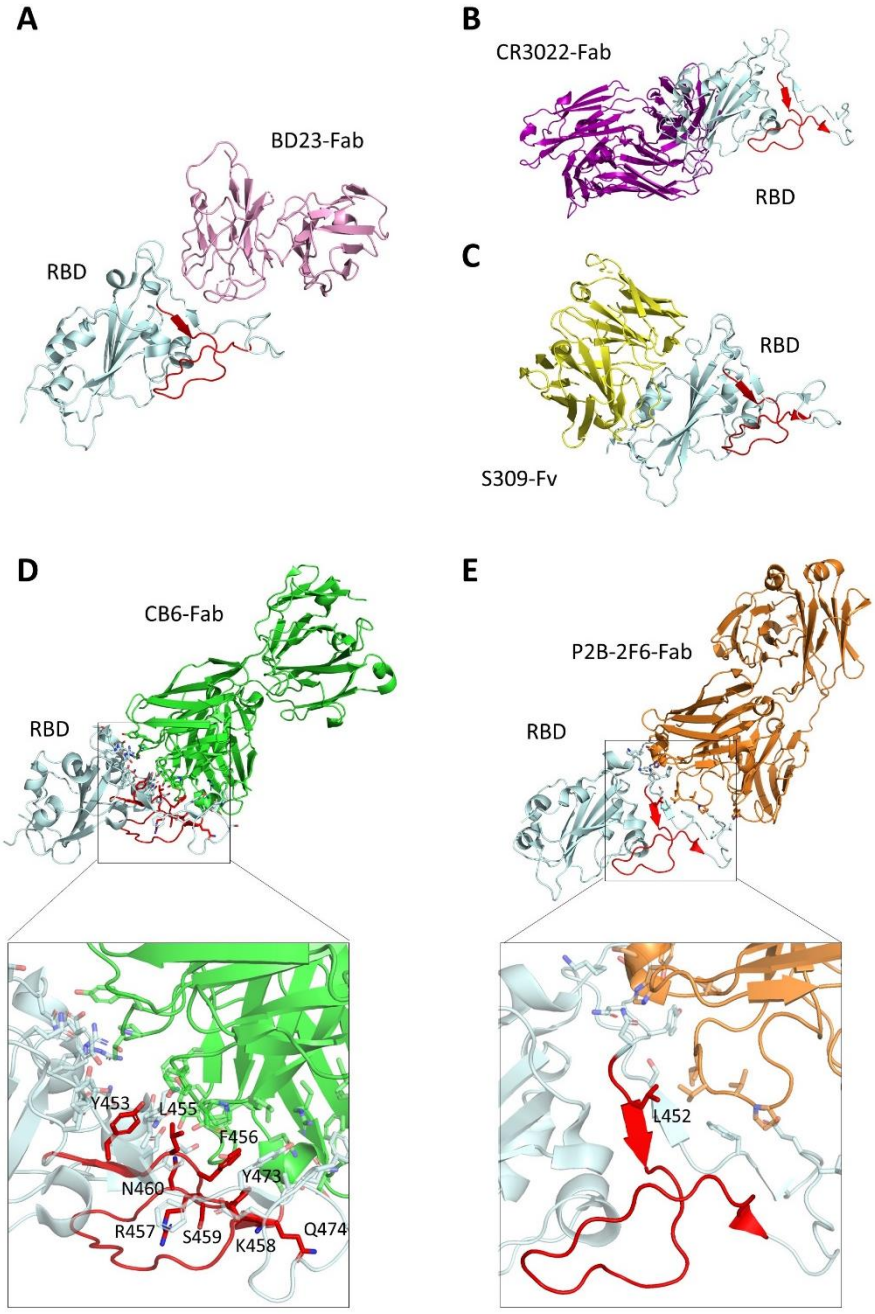


**Figure S8. The overlap of S1-76/77/78 with neutralization antibodies. A.** RBD-BD23-Fab complex (PDB: 7BYR); **B.** RBD-CR3022-Fab complex (PDB ID: 6W41); **C.** RBD-S309-Fv complex (PDB ID: 6WPT); **D.** RBD-CB6-Fab complex (PDB ID: 7C01), a close-up of the interface between RBD and CB6-Fab is depicted below, the side chain of contacting residues is shown as stick; **E.** RBD-P2B-2F6-Fab complex (PDB ID: 7BYR), a close-up of the interface between RBD and P2B-2F6-Fab is depicted below, the side chain

of contacting residues are shown as stick. The significant peptides (S1-76/77/78, aa451-474) were marked as red (**A**-**E**).

**Table. S1. Serum samples tested in this study.**

| **COVID-19 patient group** | | N = 1,051 |
|---|---|---|
| Gender | Male | 523 |
| | Female | 528 |
| Age | Mean ± s.d. | 60.3 ± 15.0 |
| | < 60 | 437 |
| | ≥ 60 | 614 |
| Severity | mild cases | 517 |
| | severe and critical cases | 534 |
| Outcome | Recovery | 972 |
| | Death | 79 |
| Sampling after symptom onset (days) | | 28.9 ± 11.2 |
| **Control group** | | N = 528 |
| Health controls | | 142 |
| Infectious diseases | | 141 |
| Lung cancer | | 48 |
| Autoimmune diseases | | 120 |
| others | | 77 |
| Gender | Male | 252 |
| | Female | 276 |
| Age | Mean ± s.d. | 53.0 ± 20.5 |
| | < 60 | 297 |
| | ≥ 60 | 231 |

## Table S2. Highly immunogenic peptides.

| NO. | Peptide ID | Sart Position | Amino acid sequence | End Position | Response frequency | Mean s.d. of control group | Mean s.d. of patient group | Note |
|---|---|---|---|---|---|---|---|---|
| 1 | S1-1 | 1 | MFVFLVLLPLVS | 12 | | | | N/A |
| 2 | S1-2 | 7 | LLPLVSSQCVNL | 18 | 12.5% | 52.7±100.2 | 133.6±220 | |
| 3 | S1-3 | 13 | SQCVNLTTRTQL | 24 | | | | N/A |
| 4 | S1-4 | 19 | TTRTQLPPAYTN | 30 | 0.3% | 14.1±28.4 | 11.4±51.3 | |
| 5 | S1-5 | 25 | PPAYTNSFTRGV | 36 | 10.8% | 83±497.1 | 1834.7±7770.5 | |
| 6 | S1-6 | 31 | SFTRGVYYPDKV | 42 | 1.0% | 19.2±85.9 | 28.3±117.5 | |
| 7 | S1-7 | 37 | YYPDKVFRSSVL | 48 | 0.2% | 18.9±35.2 | 41.8±70.7 | |
| 8 | S1-8 | 43 | FRSSVLHSTQDL | 54 | 1.3% | 44.6±88.1 | 54.6±136.8 | |
| 9 | S1-9 | 49 | HSTQDLFLPFFS | 60 | 16.1% | 59.5±105.5 | 166.6±301.5 | |
| 10 | S1-10 | 55 | FLPFFSNVTWFH | 66 | | | | N/A |
| 11 | S1-11 | 61 | NVTWFHAIHVSG | 72 | 0.0% | 3±5.9 | 11.2±8.1 | |
| 12 | S1-12 | 67 | AIHVSGTNGTKR | 78 | 1.0% | 2.8±12.8 | 12.9±84.6 | |
| 13 | S1-13 | 73 | TNGTKRFDNPVL | 84 | 0.9% | 2.4±8.4 | 12.6±69.7 | |
| 14 | S1-14 | 79 | FDNPVLPFNDGV | 90 | 0.0% | 7.5±20.4 | 11.5±25.1 | |
| 15 | S1-15 | 85 | PFNDGVYFASTE | 96 | 0.1% | 10.9±24.1 | 23.3±63.1 | |
| 16 | S1-16 | 91 | YFASTEKSNIIR | 102 | 2.0% | 31±74 | 53.8±100.4 | |
| 17 | S1-17 | 97 | KSNIIRGWIFGT | 108 | 0.0% | 2.8±4.8 | 11.1±7.3 | |
| 18 | S1-18 | 103 | GWIFGTTLDSKT | 114 | 2.0% | 46.6±84.6 | 91.2±852.3 | |
| 19 | S1-19 | 109 | TLDSKTQSLLIV | 120 | 10.8% | 69.1±137.9 | 119.9±175.7 | |
| 20 | S1-20 | 115 | QSLLIVNNATNV | 126 | 0.5% | 17.8±37.6 | 37.7±74.2 | |
| 21 | S1-21 | 121 | NNATNVVIKVCE | 132 | 0.5% | 18.9±29.4 | 69.8±71.6 | |
| 22 | S1-22 | 127 | VIKVCEFQFCND | 138 | 0.2% | 15±33 | 31.8±61.9 | |
| 23 | S1-23 | 133 | FQFCNDPFLGVY | 144 | 15.7% | 90±147.1 | 169.2±314.5 | |
| 24 | S1-24 | 139 | PFLGVYYHKNNK | 150 | 22.5% | 91.4±148.9 | 287.7±621.7 | |
| 25 | S1-25 | 145 | YHKNNKSWMESE | 156 | 0.2% | 11.5±31.2 | 30.4±65 | |
| 26 | S1-26 | 151 | SWMESEFRVYSS | 162 | 9.9% | 45±89.8 | 108.5±258.6 | |
| 27 | S1-27 | 157 | FRVYSSANNCTF | 168 | 0.2% | 8.6±16 | 73±68.8 | |
| 28 | S1-28 | 163 | ANNCTFEYVSQP | 174 | 10.9% | 45.4±82.2 | 103±185.6 | |
| 29 | S1-29 | 169 | EYVSQPFLMDLE | 180 | 0.3% | 22.4±32.8 | 48.8±69.3 | |
| 30 | S1-30 | 175 | FLMDLEGKQGNF | 186 | 1.1% | 15.6±119 | 42.4±167.1 | |
| 31 | S1-31 | 181 | GKQGNFKNLREF | 192 | 2.1% | 25.1±115.4 | 62.9±318.1 | |
| 32 | S1-32 | 187 | KNLREFVFKNID | 198 | | | | N/A |
| 33 | S1-33 | 193 | VFKNIDGYFKIY | 204 | | | | N/A |
| 34 | S1-34 | 199 | GYFKIYSKHTPI | 210 | 8.8% | 31.5±62.6 | 140.7±166.4 | |
| 35 | S1-35 | 205 | SKHTPINLVRDL | 216 | | | | N/A |
| 36 | S1-36 | 211 | NLVRDLPQGFSA | 222 | 1.0% | 6.3±16.9 | 179.9±3012.4 | |
| 37 | S1-37 | 217 | PQGFSALEPLVD | 228 | | | | N/A |
| 38 | S1-38 | 223 | LEPLVDLPIGIN | 234 | 0.2% | 11.8±28.6 | 14±36 | |
| 39 | S1-39 | 229 | LPIGINITRFQT | 240 | 0.0% | 6.7±9.2 | 33±28.6 | |
| 40 | S1-40 | 235 | ITRFQTLLALHR | 246 | | | | N/A |
| 41 | S1-41 | 241 | LLALHRSYLTPG | 252 | 0.0% | 3.6±9.7 | 9.2±23.6 | |
| 42 | S1-42 | 247 | SYLTPGDSSSGW | 258 | 0.8% | 12±41.3 | 35.2±85.4 | |
| 43 | S1-43 | 253 | DSSSGWTAGAAA | 264 | 0.0% | 5.4±13.9 | 14.6±34.1 | |
| 44 | S1-44 | 259 | TAGAAAYYVGYL | 270 | 0.5% | 10.3±103.6 | 11.7±50.3 | |
| 45 | S1-45 | 265 | YYVGYLQPRTFL | 276 | 23.2% | 26±45.5 | 250.9±264.8 | |
| 46 | S1-46 | 271 | QPRTFLLKYNEN | 282 | | | | N/A |
| 47 | S1-47 | 277 | LKYNENGTITDA | 288 | 0.8% | 13.1±122.4 | 23.5±124.3 | |
| 48 | S1-48 | 283 | GTITDAVDCALD | 294 | 0.0% | 10.3±24 | 23.9±56.9 | |
| 49 | S1-49 | 289 | VDCALDPLSETK | 300 | 0.0% | 7.2±15.9 | 17.4±43.1 | |
| 50 | S1-50 | 295 | PLSETKCTLKSF | 306 | 0.4% | 6.5±18.1 | 17.7±74.1 | |
| 51 | S1-51 | 301 | CTLKSFTVEKGI | 312 | 3.5% | 24.6±42.7 | 102.5±809.5 | |
| 52 | S1-52 | 307 | TVEKGIYQTSNF | 318 | 13.0% | 39.5±95.5 | 198.6±691.5 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 53 | S1-53 | 313 | YQTSNFRVQPTE | 324 | 0.4% | 4.7±10.6 | 21±111.8 | |
| 54 | S1-54 | 319 | RVQPTESIVRFP | 330 | 0.3% | 5±11.6 | 16.9±47.9 | |
| 55 | S1-55 | 325 | SIVRFPNITNLC | 336 | 0.6% | 16.1±29.5 | 42.3±107.3 | |
| 56 | S1-56 | 331 | NITNLCPFGEVF | 342 | 12.6% | 47.4±88.8 | 103.6±202.5 | |
| 57 | S1-57 | 337 | PFGEVFNATRFA | 348 | 22.7% | 98.2±138.7 | 280.3±287.2 | |
| 58 | S1-58 | 343 | NATRFASVYAWN | 354 | 0.0% | 11.1±13.6 | 36.3±25.4 | |
| 59 | S1-59 | 349 | SVYAWNRKRISN | 360 | 0.9% | 6.9±17.2 | 83.1±840.2 | |
| 60 | S1-60 | 355 | RKRISNCVADYS | 366 | 7.6% | 15.9±38.6 | 102±149.6 | |
| 61 | S1-61 | 361 | CVADYSVLYNSA | 372 | | | | N/A |
| 62 | S1-62 | 367 | VLYNSASFSTFK | 378 | 0.0% | 6.5±12.9 | 20.9±17.8 | |
| 63 | S1-63 | 373 | SFSTFKCYGVSP | 384 | 16.5% | 52.5±87.2 | 215±305 | |
| 64 | S1-64 | 379 | CYGVSPTKLNDL | 390 | 0.0% | 17.8±74.1 | 22.1±45.5 | |
| 65 | S1-65 | 385 | TKLNDLCFTNVY | 396 | 5.5% | 5.8±21 | 122.2±144.3 | |
| 66 | S1-66 | 391 | CFTNVYADSFVI | 402 | 15.7% | 37.8±74.1 | 190.8±327.9 | |
| 67 | S1-67 | 397 | ADSFVIRGDEVR | 408 | 0.0% | 5.1±13.4 | 27.3±32.7 | |
| 68 | S1-68 | 403 | RGDEVRQIAPGQ | 414 | 0.0% | 2.5±7.4 | 3.5±8.7 | |
| 69 | S1-69 | 409 | QIAPGQTGKIAD | 420 | 2.6% | 7.2±33.8 | 51±170.4 | |
| 70 | S1-70 | 415 | TGKIADYNYKLP | 426 | 0.3% | 4.4±14.5 | 18.3±335.1 | |
| 71 | S1-71 | 421 | YNYKLPDDFTGC | 432 | 5.8% | 46.1±121.1 | 80.2±167.4 | |
| 72 | S1-72 | 427 | DDFTGCVIAWNS | 438 | 8.4% | 34.9±73 | 84.4±164.5 | |
| 73 | S1-73 | 433 | VIAWNSNNLDSK | 444 | 0.3% | 10.8±23.7 | 30.3±96.9 | |
| 74 | S1-74 | 439 | NNLDSKVGGNYN | 450 | 0.8% | 7.2±15.2 | 28.1±232.7 | |
| 75 | S1-75 | 445 | VGGNYNYLYRLF | 456 | | | | N/A |
| 76 | S1-76 | 451 | YLYRLFRKSNLK | 462 | 5.7% | 41.5±113.6 | 156.5±919.6 | |
| 77 | S1-77 | 457 | RKSNLKPFERDI | 468 | 1.2% | 5.9±16.2 | 48.5±502.8 | |
| 78 | S1-78 | 463 | PFERDISTEIYQ | 474 | 2.2% | 19.5±141.4 | 173.5±1659.6 | |
| 79 | S1-79 | 469 | STEIYQAGSTPC | 480 | 0.0% | 13.6±44.7 | 22.2±51.6 | |
| 80 | S1-80 | 475 | AGSTPCNGVEGF | 486 | 0.1% | 6.8±25.9 | 21.5±208 | |
| 81 | S1-81 | 481 | NGVEGFNCYFPL | 492 | 0.6% | 13±28.3 | 32.9±140.4 | |
| 82 | S1-82 | 487 | NCYFPLQSYGFQ | 498 | 24.7% | 94.1±127.1 | 285.5±251.4 | |
| 83 | S1-83 | 493 | QSYGFQPTNGVG | 504 | 0.1% | 5.1±27.1 | 10.7±28.8 | |
| 84 | S1-84 | 499 | PTNGVGYQPYRV | 510 | 0.6% | 11.4±38.6 | 28.4±73.3 | |
| 85 | S1-85 | 505 | YQPYRVVVLSFE | 516 | 0.0% | 6.4±35.1 | 32.4±21.1 | |
| 86 | S1-86 | 511 | VVLSFELLHAPA | 522 | 0.0% | 10.2±19.9 | 23.1±41.5 | |
| 87 | S1-87 | 517 | LLHAPATVCGPK | 528 | 0.0% | 3.3±12 | 2.2±7.6 | |
| 88 | S1-88 | 523 | TVCGPKKSTNLV | 534 | 1.0% | 166.5±851.8 | 81.5±436 | |
| 89 | S1-89 | 529 | KSTNLVKNKCVN | 540 | 4.9% | 10.2±71.5 | 390.6±3370.3 | |
| 90 | S1-90 | 535 | KNKCVNFNFNGL | 546 | 5.8% | 23.5±40 | 137±574.1 | |
| 91 | S1-91 | 541 | FNFNGLTGTGVL | 552 | 4.5% | 32.5±70.3 | 99.6±437.2 | |
| 92 | S1-92 | 547 | TGTGVLTESNKK | 558 | 2.1% | 5.9±18.6 | 57.3±610.7 | |
| 93 | S1-93 | 553 | TESNKKFLPFQQ | 564 | 41.1% | 11.8±30.1 | 724±2300.9 | |
| 94 | S1-94 | 559 | FLPFQQFGRDIA | 570 | 12.2% | 23.3±59.5 | 223.1±523.8 | |
| 95 | S1-95 | 565 | FGRDIADTTDAV | 576 | 1.4% | 7.5±28.6 | 68.8±918 | |
| 96 | S1-96 | 571 | DTTDAVRDPQTL | 582 | 3.6% | 10±67.2 | 110.6±832 | |
| 97 | S1-97 | 577 | RDPQTLEILDIT | 588 | 49.5% | 11.6±27.7 | 897.8±1863.6 | |
| 98 | S1-98 | 583 | EILDITPCSFGG | 594 | 2.6% | 17.6±51.7 | 61.6±240 | |
| 99 | S1-99 | 589 | PCSFGGVSVITP | 600 | 0.2% | 10.7±29.4 | 56.6±63.4 | |
| 100 | S1-100 | 595 | VSVITPGTNTSN | 606 | 1.1% | 49.4±351.5 | 37.9±246.8 | |
| 101 | S1-101 | 601 | GTNTSNQVAVLY | 612 | 28.2% | 51.8±141.8 | 340.4±501.2 | |
| 102 | S1-102 | 607 | QVAVLYQDVNCT | 618 | 9.6% | 27.6±67.5 | 102.7±263.9 | |
| 103 | S1-103 | 613 | QDVNCTEVPVAI | 624 | 5.2% | 32.6±69.5 | 63.2±140.3 | |
| 104 | S1-104 | 619 | EVPVAIHADQLT | 630 | 4.3% | 4±11.2 | 95.7±582.8 | |
| 105 | S1-105 | 625 | HADQLTPTWRVY | 636 | 45.2% | 16.1±41.2 | 826.6±2029 | |
| 106 | S1-106 | 631 | PTWRVYSTGSNV | 642 | 40.3% | 64.2±110.7 | 532.3±856.6 | |
| 107 | S1-107 | 637 | STGSNVFQTRAG | 648 | | | | N/A |
| 108 | S1-108 | 643 | FQTRAGCLIGAE | 654 | 13.0% | 50.6±93.8 | 121±235.7 | |
| 109 | S1-109 | 649 | CLIGAEHVNNSY | 660 | 14.4% | 49.6±96.7 | 112.9±213.6 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 110 | S1-110 | 655 | HVNNSYECDIPI | 666 | 0.0% | 12.6±26 | 24.8±34.4 | |
| 111 | S1-111 | 661 | ECDIPIGAGICA | 672 | 49.7% | 22.1±53.1 | 963.9±2858.1 | |
| 112 | S1-112 | 667 | GAGICASYQTQT | 678 | 0.7% | 16±32 | 39.4±126.7 | |
| 113 | S1-113 | 673 | SYQTQTNSPRRA | 684 | 35.4% | 42.5±77.1 | 1191.9±4357.5 | |
| 114 | S1-114 | 679 | NSPRRARGGGGS | 685 | 1.9% | 2.1±5.4 | 54.7±531 | |
| 115 | S2-1 | 686 | SVASQSIIAYTM | 697 | | | | N/A |
| 116 | S2-2 | 692 | IIAYTMSLGAEN | 703 | | | | N/A |
| 117 | S2-3 | 698 | SLGAENSVAYSN | 709 | 0.4% | 15.6±26.4 | 32.7±98.7 | |
| 118 | S2-4 | 704 | SVAYSNNSIAIP | 715 | 6.8% | 44±120.3 | 72.2±158.1 | |
| 119 | S2-5 | 710 | NSIAIPTNFTIS | 721 | 3.6% | 21.8±83.9 | 65.5±132.8 | |
| 120 | S2-6 | 716 | TNFTISVTTEIL | 727 | 5.1% | 15.8±38.5 | 93.8±129.3 | |
| 121 | S2-7 | 722 | VTTEILPVSMTK | 733 | 0.0% | 4.3±14.9 | 2.8±9.4 | |
| 122 | S2-8 | 728 | PVSMTKTSVDCT | 739 | 2.1% | 18.4±40.6 | 45.4±104.6 | |
| 123 | S2-9 | 734 | TSVDCTMYICGD | 745 | 0.0% | 4.3±12.1 | 10.7±7.8 | |
| 124 | S2-10 | 740 | MYICGDSTECSN | 751 | 1.0% | 21±99.1 | 37±85.6 | |
| 125 | S2-11 | 746 | STECSNLLLQYG | 757 | 6.8% | 4.8±12 | 154.1±137.1 | |
| 126 | S2-12 | 752 | LLLQYGSFCTQL | 763 | 15.3% | 62.8±104.2 | 138.1±226.7 | |
| 127 | S2-13 | 758 | SFCTQLNRALTG | 769 | | | | N/A |
| 128 | S2-14 | 764 | NRALTGIAVEQD | 775 | 2.6% | 27.5±63.3 | 50±119.6 | |
| 129 | S2-15 | 770 | IAVEQDKNTQEV | 781 | 16.2% | 20.8±45.4 | 384.8±1859.5 | |
| 130 | S2-16 | 776 | KNTQEVFAQVKQ | 787 | 13.3% | 81.3±296.9 | 614.9±2321.4 | |
| 131 | S2-17 | 782 | FAQVKQIYKTPP | 793 | 1.3% | 11.8±32.6 | 29.8±103.2 | |
| 132 | S2-18 | 788 | IYKTPPIKDFGG | 799 | 12.3% | 7±33.9 | 356.9±1762.2 | |
| 133 | S2-19 | 794 | IKDFGGFNFSQI | 805 | 35.4% | 47±97.4 | 547.8±1585.3 | |
| 134 | S2-20 | 800 | FNFSQILPDPSK | 811 | 6.7% | 7.6±21.8 | 175.2±1049 | |
| 135 | S2-21 | 806 | LPDPSKPSKRSF | 817 | 2.2% | 3.6±17.2 | 74.5±949.1 | |
| 136 | S2-22 | 812 | PSKRSFIEDLLF | 823 | 45.0% | 44.6±96.5 | 2199.8±6450.8 | |
| 137 | S2-23 | 818 | IEDLLFNKVTLA | 829 | 20.2% | 53.7±120.4 | 683.7±3557.1 | |
| 138 | S2-24 | 824 | NKVTLADAGFIK | 835 | 3.9% | 48.8±89.5 | 68.3±153.5 | |
| 139 | S2-25 | 830 | DAGFIKQYGDCL | 841 | 1.8% | 23±62.5 | 72.7±572.3 | |
| 140 | S2-26 | 836 | QYGDCLGDIAAR | 847 | 0.1% | 11.2±29 | 25.7±62.9 | |
| 141 | S2-27 | 842 | GDIAARDLICAQ | 853 | 12.3% | 57.4±119 | 126.5±267.2 | |
| 142 | S2-28 | 848 | DLICAQKFNGLT | 859 | 7.7% | 23±55.1 | 70.1±158.3 | |
| 143 | S2-29 | 854 | KFNGLTVLPPLL | 865 | 0.1% | 11.1±35.6 | 16.4±60.7 | |
| 144 | S2-30 | 860 | VLPPLLTDEMIA | 871 | 0.1% | 8.5±20.5 | 15.7±39.2 | |
| 145 | S2-31 | 866 | TDEMIAQYTSAL | 877 | 0.2% | 9.8±24.9 | 24.3±124.1 | |
| 146 | S2-32 | 872 | QYTSALLAGTIT | 883 | 0.2% | 6.4±17.8 | 12.9±52.3 | |
| 147 | S2-33 | 878 | LAGTITSGWTFG | 889 | 8.3% | 24.8±58.7 | 69.4±160.2 | |
| 148 | S2-34 | 884 | SGWTFGAGAALQ | 895 | 7.4% | 65.4±99.9 | 90.6±160.2 | |
| 149 | S2-35 | 890 | AGAALQIPFAMQ | 901 | 0.7% | 7.5±24.4 | 21.7±83.9 | |
| 150 | S2-36 | 896 | IPFAMQMAYRFN | 907 | 0.6% | 27.5±45 | 100.1±71.4 | |
| 151 | S2-37 | 902 | MAYRFNGIGVTQ | 913 | 7.8% | 29.3±51.6 | 90.4±143.2 | |
| 152 | S2-38 | 908 | GIGVTQNVLYEN | 919 | 11.5% | 50.9±111.5 | 102.1±177.7 | |
| 153 | S2-39 | 914 | NVLYENQKLIAN | 925 | 13.4% | 48.1±100.6 | 107.8±204.3 | |
| 154 | S2-40 | 920 | QKLIANQFNSAI | 931 | 5.9% | 22.5±64.8 | 74.4±209.1 | |
| 155 | S2-41 | 926 | QFNSAIGKIQDS | 937 | 4.3% | 92.7±426.7 | 189.5±584 | |
| 156 | S2-42 | 932 | GKIQDSLSSTAS | 943 | 0.4% | 8.7±84.1 | 10±102.6 | |
| 157 | S2-43 | 938 | LSSTASALGKLQ | 949 | 0.2% | 6.9±23.1 | 10.2±34 | |
| 158 | S2-44 | 944 | ALGKLQDVVNQN | 955 | 0.4% | 11.1±45.3 | 16.5±191.2 | |
| 159 | S2-45 | 950 | DVVNQNAQALNT | 961 | 33.2% | 17±69.7 | 324.7±296.6 | |
| 160 | S2-46 | 956 | AQALNTLVKQLS | 967 | 12.1% | 54.9±112.2 | 142.6±864.5 | |
| 161 | S2-47 | 962 | LVKQLSSNFGAI | 973 | 8.8% | 37±61.6 | 149.6±545 | |
| 162 | S2-48 | 968 | SNFGAISSVLND | 979 | 0.3% | 20.6±115.8 | 25.5±62.6 | |
| 163 | S2-49 | 974 | SSVLNDILSRLD | 985 | 12.6% | 70.4±110.6 | 153.6±306 | |
| 164 | S2-50 | 980 | ILSRLDKVEAEV | 991 | 0.6% | 11.1±50.3 | 23.1±140.1 | |
| 165 | S2-51 | 986 | KVEAEVQIDRLI | 997 | 0.2% | 5.9±13.9 | 14.7±39.7 | |
| 166 | S2-52 | 992 | QIDRLITGRLQS | 1003 | 0.9% | 16.4±35.7 | 46.1±92.9 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 167 | S2-53 | 998 | TGRLQSLQTYVT | 1009 | 15.7% | 63.3±141.6 | 131.4±222.8 | |
| 168 | S2-54 | 1004 | LQTYVTQQLIRA | 1015 | 1.5% | 17±40.2 | 81.3±860.3 | |
| 169 | S2-55 | 1010 | QQLIRAAEIRAS | 1021 | 1.0% | 25.6±58.8 | 39.6±91.8 | |
| 170 | S2-56 | 1016 | AEIRASANLAAT | 1027 | 0.0% | 11.6±19.2 | 48.8±47.5 | |
| 171 | S2-57 | 1022 | ANLAATKMSECV | 1033 | 0.1% | 3.5±16.3 | 6.8±29.3 | |
| 172 | S2-58 | 1028 | KMSECVLGQSKR | 1039 | 0.1% | 2.3±14.5 | 3.5±25.9 | |
| 173 | S2-59 | 1034 | LGQSKRVDFCGK | 1045 | 1.1% | 50.1±292.2 | 40.1±247.3 | |
| 174 | S2-60 | 1040 | VDFCGKGYHLMS | 1051 | 0.0% | 11.4±25.9 | 40.7±55.7 | |
| 175 | S2-61 | 1046 | GYHLMSFPQSAP | 1057 | 0.0% | 7±12.1 | 24.2±25.6 | |
| 176 | S2-62 | 1052 | FPQSAPHGVVFL | 1063 | 17.2% | 103.5±196.3 | 287.9±409.6 | |
| 177 | S2-63 | 1058 | HGVVFLHVTYVP | 1069 | 0.2% | 10.8±21.8 | 16.7±42.7 | |
| 178 | S2-64 | 1064 | HVTYVPAQEKNF | 1075 | 1.0% | 8.7±46.2 | 30.5±154.2 | |
| 179 | S2-65 | 1070 | AQEKNFTTAPAI | 1081 | 0.3% | 9.8±44.7 | 6.8±57.8 | |
| 180 | S2-66 | 1076 | TTAPAICHDGKA | 1087 | 0.1% | 6.9±36.8 | 3.9±19.5 | |
| 181 | S2-67 | 1082 | CHDGKAHFPREG | 1093 | 0.3% | 4.6±11.6 | 18.3±59.3 | |
| 182 | S2-68 | 1088 | HFPREGVFVSNG | 1099 | 0.1% | 10.5±23.1 | 26.5±60.5 | |
| 183 | S2-69 | 1094 | VFVSNGTHWFVT | 1105 | 14.2% | 30.1±66.1 | 110.9±221.7 | |
| 184 | S2-70 | 1100 | THWFVTQRNFYE | 1111 | 15.1% | 28.4±64.9 | 128.4±293.1 | |
| 185 | S2-71 | 1106 | QRNFYEPQIITT | 1117 | 12.2% | 74.5±176.7 | 187.5±589.4 | |
| 186 | S2-72 | 1112 | PQIITTDNTFVS | 1123 | 0.1% | 7±25.7 | 18.8±39.8 | |
| 187 | S2-73 | 1118 | DNTFVSGNCDVV | 1129 | 6.3% | 58.4±244.8 | 125±229 | |
| 188 | S2-74 | 1124 | GNCDVVIGIVNN | 1135 | | | | N/A |
| 189 | S2-75 | 1130 | IGIVNNTVYDPL | 1141 | 0.5% | 4.1±11.1 | 87.5±69.7 | |
| 190 | S2-76 | 1136 | TVYDPLQPELDS | 1147 | 1.0% | 12.3±40.4 | 35.9±121 | |
| 191 | S2-77 | 1142 | QPELDSFKEELD | 1153 | 1.2% | 9.3±23.9 | 40.1±126.9 | |
| 192 | S2-78 | 1148 | FKEELDKYFKNH | 1159 | 69.3% | 16.8±166.1 | 2173.3±4176.5 | |
| 193 | S2-79 | 1154 | KYFKNHTSPDVD | 1165 | 1.6% | 7.2±17.2 | 41.4±121 | |
| 194 | S2-80 | 1160 | TSPDVDLGDISG | 1171 | 1.1% | 6.4±18.9 | 41±301.3 | |
| 195 | S2-81 | 1166 | LGDISGINASVV | 1177 | 10.6% | 12.7±54 | 162.9±435.5 | |
| 196 | S2-82 | 1172 | INASVVNIQKEI | 1183 | 1.2% | 13.8±74.7 | 61.5±415 | |
| 197 | S2-83 | 1178 | NIQKEIDRLNEV | 1189 | 7.2% | 7.9±27.4 | 147.8±780.6 | |
| 198 | S2-84 | 1184 | DRLNEVAKNLNE | 1195 | 1.2% | 6.3±42.1 | 36.7±117.1 | |
| 199 | S2-85 | 1190 | AKNLNESLIDLQ | 1201 | 10.5% | 62.2±117 | 148.2±379 | |
| 200 | S2-86 | 1196 | SLIDLQELGKYE | 1207 | 0.0% | 90.4±1181.4 | 109.5±190.6 | |
| 201 | S2-87 | 1202 | ELGKYEQYIKWP | 1213 | 9.2% | 33.2±54 | 144.4±139.4 | |
| 202 | S2-88 | 1208 | QYIKWPWYIWLG | 1219 | 16.3% | 35.5±63.1 | 180.7±238.4 | |
| 203 | S2-89 | 1214 | WYIWLGFIAGLI | 1225 | | | | N/A |
| 204 | S2-90 | 1220 | FIAGLIAIVMVT | 1231 | | | | N/A |
| 205 | S2-91 | 1226 | AIVMVTIMLCCM | 1237 | | | | N/A |
| 206 | S2-92 | 1232 | IMLCCMTSCCSC | 1243 | | | | N/A |
| 207 | S2-93 | 1238 | TSCCSCLKGCCS | 1249 | 15.6% | 35.4±80.7 | 112.8±222.2 | |
| 208 | S2-94 | 1244 | LKGCCSCGSCCK | 1255 | | | | N/A |
| 209 | S2-95 | 1250 | CGSCCKFDEDDS | 1261 | 0.6% | 10.7±24 | 36.7±347.2 | |
| 210 | S2-96 | 1256 | FDEDDSEPVLKG | 1267 | 13.1% | 6.1±15.8 | 350.7±1430.8 | |
| 211 | S2-97 | 1262 | EPVLKGVKLHYT | 1273 | 20.3% | 9.7±103.6 | 474.6±2287.5 | |

## Table S3. Highly immunogenic peptides.

| NO. | Peptide | Sequence | Position | Domain | Response frequency | Zhang *et al.* (31) | Peptide array (22) | ReScan (23) | COVIDep (21) | GRAVY | pI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | S1-5 | PPAYTNSFTRGV | 25-36 | S1-NTD | 10.75% | 21-45 | 26-30 | - | - | -0.53 | 9.84 |
| 2 | S1-93 | TESNKKFLPFQQ | 553-564 | S1-CTD1 | 41.10% | 522-646 | - | 552-589 | 563-568 | -1.29 | 9.88 |
| 3 | S1-97 | RDPQTLEILDIT | 577-588 | S1-CTD1 | 49.48% | 522-646 | - | 552-589 | - | -0.41 | 3.71 |
| 4 | S1-105 | HADQLTPTWRVY | 625-636 | S1-CTD2 | 45.20% | 522-646 | - | - | - | -0.84 | 7.88 |
| 5 | S1-106 | PTWRVYSTGSNV | 631-642 | S1-CTD2 | 40.34% | 522-646 | - | - | - | -0.57 | 9.84 |
| 6 | S1-111 | ECDIPIGAGICA | 661-672 | S1-CTD2 | 49.67% | - | - | - | 667-674 | 1.06 | 3.01 |
| 7 | S1-113 | SYQTQTNSPRRA | 673-684 | S1-CTD2 | 35.39% | - | - | - | - | -1.97 | 11.16 |
| 8 | S2-15 | IAVEQDKNTQEV | 770-781 | S2 | 16.18% | - | - | - | - | -0.91 | 3.83 |
| 9 | S2-16 | KNTQEVFAQVKQ | 776-787 | S2 | 13.32% | - | - | - | - | -1.08 | 9.88 |
| 10 | S2-18 | IYKTPPIKDFGG | 788-799 | S2 | 12.27% | - | - | 799-836 | - | -0.46 | 9.63 |
| 11 | S2-19 | IKDFGGFNFSQI | 794-805 | S2 | 35.39% | - | - | 799-836 | - | 0.12 | 7.00 |
| 12 | S2-22 | PSKRSFIEDLLF | 812-823 | S2-FP | 45.00% | - | 806-820 | 799-836 | 818-833 | -0.08 | 7.00 |
| 13 | S2-23 | IEDLLFNKVTLA | 818-829 | S2-FP | 20.17% | - | - | 799-836 | 818-833 | 0.80 | 4.07 |
| 14 | S2-78 | FKEELDKYFKNH | 1148-1159 | S2 | 69.27% | - | - | 1141-1178 | 1136-1152, 1157-1173 | -1.73 | 7.66 |
| 15 | S2-96 | FDEDDSEPVLKG | 1256-1267 | S2-CT | 13.13% | - | - | - | 1255-1273 | -1.12 | 3.43 |
| 16 | S2-97 | EPVLKGVKLHYT | 1262-1273 | S2-CT | 20.27% | - | - | - | 1255-1273 | -0.21 | 9.63 |
| 17 | S1-76 | YLYRLFRKSNLK | 451-462 | S1-RBD | 5.71% | 450-469 | 456-460 | - | - | -0.79 | 10.78 |
| 18 | S1-77 | RKSNLKPFERDI | 457-468 | S1-RBD | 1.24% | 450-469 | 457-460 | - | - | -1.55 | 10.57 |
| 19 | S1-78 | PFERDISTEIYQ | 463-474 | S1-RBD | 2.19% | 450-469 | - | - | - | -0.93 | 3.83 |